BMJ Open

Analyzing Counterintuitive Data

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-026447
Article Type:	Research
Date Submitted by the Author:	02-Sep-2018
Complete List of Authors:	Celi, Leo Anthony; Massachusetts Institute of Technology, Doty, Erik; Harvard Medical School Department of Biomedical Informatics Stone, David; University of Virginia School of Medicine; Massachusetts Institute of Technology, Laboratory for Computational Physiology McCague, Ned; Massachusetts Institute of Technology, Laboratory for Computational Physiology; Kyruus, Inc
Keywords:	Pain, Mortality, Length of Stay

Analyzing Counterintuitive Data

Erik Doty[1], DO, MBI; David J. Stone[2,5], MD; Ned McCague[3,5], MPH; Leo Anthony Celi[4,5], MD, MS, MPH

Corresponding author: Leo Anthony Celi

Email: lceli@mit.edu

Phone: (617) 253-7937

Word Count: 3711

1. *Baystate Medical Center, Springfield, MA*
2. *University of Virginia School of Medicine, Charlottesville, VA*
3. *Kyruus, Inc., Boston, MA*
4. *Beth Israel Deaconess Medical Center, Boston, MA*
5. *MIT Laboratory for Computational Physiology, Cambridge, MA*

Analyzing counterintuitive data

E. Doty, D.J. Stone, N. McCague, L.A. Celi

Keywords: Pain, mortality, length of stay,

Abstract

Objective: To explore the issue of counterintuitive data via analysis of a representative case, in which we explore the relationship between perceived pain in the ICU and patient outcomes of interest, with further discussion of situations in which the data appear to be inconsistent with current knowledge.

Design: Retrospective analysis of a cohort of CABG patients derived from the MIMIC-III database. Regression analysis was used to examine the association between perceived pain in the ICU and patient outcomes.

Setting: MIMIC-III database, a publicly available, deidentified critical care patient database.

Participants: 844 patients were selected from the database that met the following inclusion criteria: Adult > 18 years old, underwent CABG surgery, and extubated within 24 hours after ICU admission; and no exclusion criteria: Non-CABG surgery and missing data on confounding variable.

Outcomes: 30 Day mortality, 1-year mortality, and hospital length of stay.

Results: Increased levels of pain were found to be significantly associated with reduced mortality at 30 days and 1-year, and shorter hospital LOS. A one-point increase in mean pain level was found to be associated with a reduction in the odds of 30-day and 1-year mortality by a factor of 0.457 (95%CI 0.304-0.687, p< 0.01) and 0.710 (95%CI 0.571 - 0.881, p< 0.01) respectively, and a 0.916 (p < 0.01) day decrease in hospital LOS.

Conclusion: The reliability of counterintuitive results must be particularly carefully examined. We suggest several issues to consider in this process. If the data is determined, so far as possible, to be valid, consideration must then be made towards alternative explanations for the unexpected results observed. Such results may in fact indicate that current knowledge is incomplete and function to inspire further research into the factors involved.

Strengths and limitations of this study

- Large sample size with complete covariate data.
- Multiple regression models with multiple sensitivity analyses.
- High internal validity shown by use of falsification hypothesis testing.
- Lack of oral analgesic data.
- Recognizing that correlation does not equal causation and further work is needed to confirm case results.

Introduction

What do we mean by counterintuitive data? It is data that presents unexpected results that may

clash with common sense or what has been previously published and accepted by the medical community. In practice, clinicians have long dealt with such results in individual bits but have had the vast advantage of being able to examine the concurrent state of the patient and react in real time by repeating a lab test or tracking ongoing monitor data. These responses function to identify the prior result as a non-repeatable error, or as a genuine anomaly. However, this approach is not applicable to the context of retrospective data analysis. Furthermore, the counterintuitive data revealed in such analyses is likely to be more involved than a single aberrant lab or vital sign value. In today's data driven healthcare system, retrospective data analyses are becoming more and more common. We can therefore logically expect to encounter a greater incidence and variety of counterintuitive values and results that are impossible to confirm by repetition, difficult to confirm or deny by context, but still require interpretation.

The question then becomes how best to approach such results? Are they incorrect simply because they weren't what was expected? And was the expectation itself based on subjective assumptions or objective conclusions? When our prior expectations are not met, are we dealing with truly faulty data, or do our expectations need to be reset by what are reliable, but counterintuitive, results. For example, we have learned that intensive care practices common in the past such as large tidal volume ventilation, the use of pulmonary artery catheters, and the use of lidocaine infusions in myocardial infarction led to no benefit or injury.[1-3] Were these unexpected negative outcomes initially missed because outcomes data was not being carefully analyzed, or perhaps ignored or interpreted as counterintuitive to the level of unbelievability? How can the situation be dissected retrospectively so that counterintuitive data can be identified as truly spurious versus simply not being consistent with our prior experience which may itself be faulty and require data driven correction?

In this paper, we explore a case in which the results contradicted previous reports and our clinical expectations. Using the Medical Information Mart for Intensive Care-III (MIMIC-III), a critical care database that was developed and maintained by the Laboratory for Computational Physiology at the Massachusetts Institute of Technology[4], we retrospectively selected a cohort of patients that underwent a coronary artery bypass graft (CABG) procedure and evaluated the effect of perceived pain on mortality and hospital length of stay (LOS). Our initial hypothesis

was that increased levels of perceived pain would correlate with worse patient outcomes such as increased hospital length of stay. This would be in line with the current literature that suggest optimal pain control leads to increased mobility, earlier ambulation, and improved outcomes. [5-7] Contrary to the literature, we found that higher levels of pain were associated with reduced mortality and reduced LOS. We then discuss potential causes of these results and the general issue of dealing with counterintuitive results in retrospective data analyses.

Case

Population

We selected patients from the MIMIC database who met all of the following inclusion criteria and none of the exclusion criteria. Inclusion criteria included: (1) Adult > 18 years old, (2) who underwent CABG surgery, and (3) was extubated within 24 hours after arrival to the ICU. Exclusion criteria were: (1) Non-CABG surgical procedure, and (2) missing data on confounding variables. Patients were identified using Current Procedural Terminology (CPT) codes: The following CPT codes corresponded to the CABG procedure: 33510 to 33516 for venous grafting only for coronary artery bypass, and 33533 to 33548 for arterial grafting for coronary bypass. The final study cohort contained 844 patients (***Figure 1***).

The MIMC-III database included 1,917 patients who underwent CABG, with 844 meeting the study criteria. CABG was chosen for the investigation as it is a common procedure with the majority of patients having no or few post-operative complications and relatively predictable recoveries.[5] Due to the nature of the surgical procedure which requires sternal spreading for exposure, there is an expected high analgesic burden immediately after surgery.

Outcomes

The primary outcome assessed was mortality at 30 days. Secondary outcomes were mortality at 1 year and hospital LOS. In the MIMIC database, mortality data for patients who die after hospital discharge is derived from the social security death registry.[4]

Exposures

The exposures of interest were pain levels reported by the patient immediately and in the subsequent interval after ICU extubation. Pain levels on a scale of 0-10 were regularly self-

reported by patients to ICU nurses and recorded in the database, generating a continuum of measurements for each patient. The mean, median, and maximum pain levels were used for separate analyses. Concomitant measurements of heart rates, respiratory rates, and systolic blood pressures were also compared against their simultaneously recorded pain measurement.

Intravenous (IV) opiate administration was extracted from the database. MIMIC contained data for the following medications: Morphine, fentanyl, hydromorphone, and meperidine. The was no data in MIMIC corresponding to the administration of oral analgesics.

Nausea and delirium were also tested against our outcomes. The presence of nausea was derived from the nursing notes stored in the database. A positive nausea exposure was defined as the mention "nausea" or "nauseous" in the nursing note with no negative descriptor, such as "not nauseous" or "denies nausea", attached. Delirium was similarly assessed by looking for mention of "delirium", "delirious", or "confusion". Additionally, delirium exposure was considered positive if patients had a positive nursing delirium assessment.

Covariates

Several variables found to be linked to worse patient outcomes in previous studies were included to control for confounding in the regression models: demographic factors, comorbid conditions, and illness severity score on admission to the ICU.[8,9] Comorbid burden was represented by the Elixhauser index which is determined by the aggregate presence or absence of 30 different comorbid conditions as detected by ICD-9 codes.[10] Illness severity was captured using the Oxford Acute Severity of Illness Score (OASIS), which is calculated on admission to the ICU and takes into account age, heart rate, Glasgow coma scale, mean arterial pressure, temperature, respiratory rate, ventilatory status, urine output, pre-ICU in-hospital LOS, and whether or not the patient underwent elective surgery. Studies have shown OASIS is comparable to other illness severity ratings in predicting outcomes such as mortality and length of stay.[11]

Analysis

Analysis was carried out using R version 3.4.0 and SAS 9.4. Unconditional logistic regression with Fisher's optimization was used to compare the pain measures with 30-day and 1-year mortality. Linear regression was used to model the relationship between mean pain scores and hospital LOS. Age, gender (male reference), Elixhauser index, and OASIS score were included

in the models to account for potential confounders. In a separate ordinal regression, mean pain levels were categorized into four groups of no pain (0/10), mild pain (1-3), moderate pain (3-6), and severe pain (7-10) in accordance with the NIH Pain Consortium.[12]

ANOVA was used to determine if there was a significant variation in heart rate, respiratory rate, and/or systolic blood pressure, when compared to the concurrent pain assessment.

IV analgesia medications were converted to their morphine equivalents based on the National Pharmaceutical Counsel's guidelines.[13] The IV analgesia was subdivided into total dose in the first 24 hours, mean dose per ICU course day, and total dose during ICU course. ANOVA models were used to determine if there were any significant variation in administration of IV analgesics among the four categorized pain groups.

Two sensitivity analyses were performed to assess the robustness of the observed effects. The first included the same statistical tests in all postoperative CABG patients regardless of duration of intubation. The second sensitivity analysis excluded patients who died in the hospital. To add validity to the potential associations, falsification hypothesis testing using nausea, a symptom with no known effect on clinical outcomes, was performed on the same patient cohort. Assessment of delirium, a symptom associated with poorer patient outcomes, was also performed against the outcome measures.[14]

Results

The database included 844 patients who underwent a CABG procedure and were extubated within 24 hours. There were 68 patients who on average reported no pain during their ICU stay after extubation, 419 with mild pain, 336 with moderate pain, and 21 with severe pain. The distribution of patient characteristics, including age, gender, illness acuity on ICU admission (OASIS), and comorbidity index is reported in ***Table 1***. There was no significant difference noted in the frequency in which pain was assessed in those who experienced lower pain levels when compared to those who experienced increased pain levels. The number of comorbidities ranged from 0 to 9. Bivariate analysis showed increasing OASIS was significantly associated with increased mortality and increased LOS ($p < 0.05$). No significant differences were found in the amount of IV analgesia administered among the pain subgroups.

Bivariate analysis (***Figure 2***) shows a correlation between increasing pain levels and improved

outcomes among these patients who had no intra-operative complications and were extubated within 24 hours of arrival in the ICU. Higher pain levels for this specific cohort of patients who were fast-tracked after CABG were found to be associated with decreased hospital LOS. Those who experienced lower levels of pain in the ICU were more likely to be dead at 30 days and 1 year.

Multivariate regression analysis was performed to adjust for confounding. Four different models using mean, median, and maximum pain scores, and pain categories were tested against the clinical outcomes with the results displayed in ***Table 2***. The logistic regression models consistently showed that increasing pain was associated with reduced odds of death at 30 days and 1 year after adjustment for illness severity and co-morbid conditions. All the linear models demonstrated that increasing pain levels were also associated with decreased hospital LOS, except for the model that looked at the maximum pain score, which showed an opposite effect.

No significant variations were noted in heart rate, respiratory rate, or blood pressure with increasing pain levels.

Sensitivity analysis was employed to examine all patients regardless of duration of intubation, expanding the sample size to 1889 patients. The results were similar for 30-day mortality and hospital LOS as regards effect size and statistical significance; however, the results were not statistically significant for 1-year mortality (***Table 2***). An additional sensitivity analysis excluded patients who died in the hospital- these results were consistent with the prior models and were statistically significant for hospital LOS, but not for mortality (***Table 2***).

As expected, the presence of nausea was not found to be associated with any impact on outcomes in the study cohort. As also would be expected, patients who had delirium had worse 30-day and 1-year mortality and longer hospital LOS.

Discussion

We will first discuss our unexpected results, and then discuss the general issue of counterintuitive data. Our results that increasing levels of patient-reported pain severity post-CABG surgery are associated with better clinical outcomes were not consistent with our initial hypothesis that better outcomes would correlate with better pain control as per the reported

literature. In fact, prior studies have found increased levels of pain in the hospital to be associated with increased mortality. [15]

The difference in the study cohort between our study and others may explain some of the discordance. Our initial analysis was limited to "fast-tracked" patients who did not have intra-operative complications and were extubated early in their ICU course. These patients made up 44% of the database patients. Studies that have reported worse clinical outcomes associated with post-operative pain did not select for a relatively healthy sub-cohort of patients. Why would patients with higher levels of pain have better outcomes? It is well documented that an increased inflammatory reaction is associated with increased pain. Pro-inflammatory cytokines such as IL-1β, IL-6, and TNF-α have been directly implicated in the physiology of pain.[16,17] These cytokines have also been found to be directly involved in wound healing through the stimulation of processes such as keratinocyte and fibroblast proliferation, and synthesis and breakdown of extracellular matrix proteins.[18] We speculate that those patients who demonstrated better outcomes mounted a more robust inflammatory response leading to more pain, but also to increased healing ability.

Another possibility is that higher perceived pain levels represent a proxy for a generally better state of health, including superior physiological function of the cardiovascular, respiratory, renal, and hepatic systems. In tandem, these systems act to metabolize and eliminate anesthetic and analgesic drugs so that the net pharmacokinetic result would likely be increased susceptibility to pain due to less administered agent remaining at active sites. Furthermore, patients with better cardiovascular function would likely have better cerebral perfusion with improved central neurological function, and thereby have a pharmacodynamic reason for perceiving more pain. And patients who are generally in better overall condition would be expected to manifest better outcomes. These thoughts are admittedly speculative and additional research is needed to explore these possibilities.

It is important to point out that the goal of clinicians should not be in any way to maximize pain to optimize outcomes. Conventional approaches that aim to control pain adequately should be employed. Our observation is just that - an observation of an association and conjectures of possible linking mechanisms but is not intended in any way to drive pain management policy in the direction of tolerating undertreated pain.

We performed sensitivity analyses, one including all patients regardless of post-operative ventilation duration, and another excluding patients who died during hospitalization, and reached similar conclusions. When excluding in-hospital deaths, we discovered the 30-day mortality rate had a similar odds ratio but was no longer statistically significant. This is most likely due to the low mortality rate after hospital discharge following CABG, making it difficult to detect a statistically significant effect.

We believe that researcher bias is a non-issue as these findings were not expected, but rather, the opposite. Sampling bias was also minimal. Our inclusion criteria were predefined prior to database sampling. We performed multiple sensitivity analyses to determine if those that were excluded would have had an effect on our results. However, the study has several limitations inherent in any retrospective data analysis. We acknowledge that correlation does not equal causation and further research is needed to determine the underlying physiologic mechanism for the results seen. Due to the self-reported nature of the pain scores, reporting bias is a concern. Some patients may have over-reported and others under-reported their pain. We also recognize that analgesic administration is a confounder. While we were unable to directly control for this due to lack of information regarding anesthetic and pain management in the database, we attempted to limit this potential confounder by excluding those with prolonged intubations who would inherently have received and required greater doses of sedatives and analgesics. Despite measures taken to guarantee internal validity, we anticipate appropriate skepticism with regard to generalizability of the findings. This, of course, is of genuine concern given the current state-of-affairs where clinicians are already inundated with conflicting studies of questionable quality. We therefore invite other investigators to replicate (and expand) our analysis in other databases.

As noted, our findings were contrary to clinical expectations and to most published works which associate increased pain with worse outcomes.[15,19-20] Encountering counterintuitive results is not unique to retrospective data analysis. Clinicians encounter unexpected, possibly aberrant, values in situations such as the evaluation of laboratory and monitor data. When a possibly spurious lab result is obtained, the usual response is to repeat the test. When the second test comes back with a more acceptable value, we generally then ignore the unexpected value. But what if the repeat value is also aberrant? Do we repeat it again, or do we begin to believe that the value is 'real' and start to formulate a response to a clinical problem? In this case, it is the *consistency* and

reproducibility of the counterintuitive value that drives its possible validity. The details of this process are determined by the overall clinical risks involved. The consistency we found in the pain score values drove us to consider the possibility that the values were 'real' even though they were counterintuitive in terms of our expectations.

Another issue in evaluating to counterintuitive values is whether they are *possible.* Impossible values would include a potassium of 64.5, one incompatible with life. But a potassium of 7.3 is a possible value. The pain values associated with better outcomes were unexpected, but not so high that they were impossible in the observed context.

One question that would arise with a potassium of 7.3 would be that of continuity- did the value occur suddenly or gradually in a stream of normal values? Were surrounding values similarly abnormal? In the context of persistently abnormal values, e.g. untreated uremia, a normal value would be counterintuitive. So that while most counterintuitive values will tend to be out of the 'normal range', they will not necessarily be so. In the context of increasing values, it might simply be the first one that was not only out of the normal range, but that crossed the line into a critical range,

The fundamental question is whether counterintuitive results are actually false results, or does the problem lie in our perception of what should be. Table Three displays a categorization of error types that could result in faulty data. We are not able to attribute the counterintuitive data we observed to any of these factors, however.

How can counterintuitive results be approached in secondary data analyses? Table Four displays characteristics that may distinguish reliable (but counterintuitive) from truly faulty data. With consideration of these factors, the first investigative step is to retrace the process and workflow involved in data entry so far as possible. Our data was obtained at the institution of several of the authors where nurses are trained to assess pain on a standard scale from 0 to 10. There are several potential faults to this method. The nursing staff could neglect to regularly assess pain or neglect to enter the information into the medical record generating the database. While this may alter a few data points, it is unlikely to systematically affect all data unless there was an obvious glaring institutional issue affecting every nurse and every data entry.

After determining that the data source is valid, additional statistical tests can be run on the patient cohort. Tests such as the falsification hypothesis testing we utilized, add validity to the results as they show that the cohort follows other generally known principles. In our study, falsification analysis by both neutral (nausea) and positive (delirium) factors provided support for our findings.

Concurrent contextual data can also help to confirm the veracity of data- for example, one could examine ECGs if hyperkalemia was being analyzed. We examined concomitant vital signs during the time of pain measurements. We expected to observe significant increases with higher pain levels, but did not: With the combination of analgesics, residual anesthetics, and the concurrent use of drugs that directly affect vital signs such as beta-blockers, the lack of correlation is probably not surprising. In fact, we learned that in this setting, it appears to be inadvisable to use vital sign changes as a proxy for the presence of unvoiced pain. Finally, one can attempt to physiologically explain the disparity between the observed and expected results as we did above for the case of post- CABG pain.

The use of lower thresholds for blood transfusions in the ICU is an example of a counterintuitive finding. ICU target hemoglobin levels were historically set at greater than 10 g/dL, theoretically to ensure adequate oxygen delivery.[21] This led to increased transmission of blood borne diseases, unnecessary healthcare expenditures, and actually worse outcomes.[22] Later research showed that this rule was not necessary for most patients, but only for selected patients such as those with acute coronary syndrome actively experiencing chest pain. The initially counterintuitive findings that lower hemoglobin levels were not only acceptable but preferable in most cases, served as research triggers to more fully elucidate optimal clinical practice. Our case may serve as an analogous research trigger in terms of optimally managing postoperative pain. Outcomes such as mortality and LOS are complex phenomena driven by many factors- to observe a clear and robust statistical effect such as we did is strongly suggestive that something 'real' is occurring even if the data were counterintuitive.

The final step when dealing with counterintuitive data is to look for additional evidence that confirms the reliability of the results (perhaps this could be termed 'confirmatory metadata'). With respect to our CABG case, the analysis should be rerun on additional databases and in different settings. Just as clinicians continued to manage intensive care unit anemia as they

always had until more definitive results were reported, our results should not impact the analgesic care of patients at this point. However, we hope that we have raised the issue in the appropriate minds that outcomes may benefit from approaches slightly different from usual. After all, one can easily eliminate all pain from postoperative patients but they would have to remain sedated and ventilated for an indefinite period of time to do so. And after they are extubated, pain management should not be so aggressive that it leads to apnea and respiratory arrest. In other words, there may be a detectable level of tolerable pain that leads patients to their best outcomes, and no honest clinician will guarantee a patient that they will have no pain at all after a procedure like a sternal-disrupting CABG.

Conclusion

Contrary to our expectations, we observed, in a retrospective analysis of electronic health records, that post-CABG fast-track patients with higher pain scores had better outcomes. The increasing use of EHRs for secondary analysis will likely lead to an increasing incidence of such apparently counterintuitive results. While the first step in this situation is to attempt to confirm the reliability of both the analytic process and the data itself, such findings that prove to be robust may lead to further ideas and subsequent research that drive future clinical care. On the other hand, clinicians must be careful in terms of modifying their practices until the implications of such counterintuitive (or any) data have been thoroughly vetted and confirmed in diverse database contexts and via the peer review process.

Declarations

Availability of Data Materials

The datasets generated for the current study were derived from the MIMIC-III Database available at https://mimic.physionet.org/. The data subsets and statistical code used in this project can be found at https://github.com/ErikWDoty/PainProject

Consent for Publication

Not applicable

Competing Interests

The authors declare they have no completing interests.

Funding

There was no financial support for this project.

Author Contributions

ED was responsible for the data extraction, the initial statistical analysis, and writing and editing the manuscript. NM was involved in validating the statistical models and participated in editing the manuscript. DS was responsible for assisting with background information and editing the manuscript. LC was the project supervisor, responsible for project conception and manuscript editing.

Acknowledgements

We would like to thank J. Michael Jaeger, MD, PhD of the University of Virginia School of Medicine, for his assistance with background information.

References

[1] Gajic O, Dara SI, Mendez JL, et al. Ventilator-associated lung injury in patients without acute lung injury at the onset of mechanical ventilation. Critical Care Medicine. 2004;32:1817-1824.

[2] A.F Connors Jr, T Speroff, N.V Dawson, et al. The effectiveness of right heart catheterization in the initial care of critically ill patients (SUPPORT Investigators) JAMA, 276 (1996), pp. 889-89

[3] Hine LK, Laird N, Hewitt P, Chalmers TC. Meta-analytic Evidence Against Prophylactic Use of Lidocaine in Acute Myocardial Infarction. Arch Intern Med. 1989;149(12):2694–2698. doi:10.1001/archinte.1989.00390120056011

[4] MIMIC-III, a freely accessible critical care database. Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. Scientific Data (2016). DOI: 10.1038/sdata.2016.35. Available from: http://www.nature.com/articles/sdata201635

[5] Sutherland, J., Liu, G., Crump, T., Bair, M., & Karimuddin, A. (2018). Relationship between preoperative patient-reported outcomes and hospital length of stay: A prospective cohort study of general surgery patients in Vancouver, Canada. *Journal of Health Services Research & Policy*. doi:10.1177/1355819618791634

[6] Bykowski, M. R., Sivak, W., Garland, C., Cladis, F. P., Goldstein, J. A., & Losee, J. E. (2018). A Multimodal Preemptive Analgesic Protocol for Alveolar Bone Graft Surgery: Decreased Pain, Hospital Stay, and Health Care Costs. *The Cleft Palate-Craniofacial Journal*. doi:10.1177/1055665618791943

[7] Ljungqvist O, Scott M, Fearon KC. (2017) Enhanced Recovery After Surgery: A

Review. *JAMA Surg*;152(3):292–298. doi:10.1001/jamasurg.2016.4952

[8] Hawkes, A. L., Nowak, M., Bidstrup, B., & Speare, R. (2006). Outcomes of coronary artery bypass graft surgery. Vascular Health and Risk Management, 2(4), 477–484.Hawkes, A. L., Nowak, M., Bidstrup, B., & Speare, R. (2006). Outcomes of coronary artery bypass graft surgery. *Vascular Health and Risk Management*, *2*(4), 477–484.

[9] Shroyer, A. W., Coombs, L. P., Peterson, E. D., Eiken, M. C., Delong, E. R., Chen, A., Edwards, F. H. (2003). The society of thoracic surgeons: 30-day operative mortality and morbidity risk models. The Annals of Thoracic Surgery, 75(6), 1856-1865. doi:10.1016/s0003-4975(03)00179-6

[10] Elixhauser, A., Steiner, C., Harris, D., & Coffey, R. (1998). Comorbidity Measures for Use with Administrative Data. Medical Care, 36(1), 8-27. Retrieved from http://www.jstor.org/stable/3766985

[11] Johnson, A. E., Kramer, A. A., & Clifford, G. D. (2013). A New Severity of Illness Scale Using a Subset of Acute Physiology and Chronic Health Evaluation Data Elements Shows Comparable Predictive Accuracy*. Critical Care Medicine, 41(7), 1711-1718. doi:10.1097/ccm.0b013e31828a24fe

[12] Pain Intensity Instruments. (2003, July). Retrieved April 9, 2017, from https://painconsortium.nih.gov/pain_scales/NumericRatingScale.pdf

[13] *Pain: Current understanding of assessment, management, and treatments*. (2001). Reston, VA: National Pharmaceutical Council.

[14] Prasad, V., & Jena, A. B. (2013). Prespecified Falsification End Points. *Jama, 309*(3), 241. doi:10.1001/jama.2012.96867

[15] Lewis KS, Whipple JK, Michael KA, Quebbeman EJ. Effect of analgesic treatment on

the physiological consequences of acute pain. Am J Hosp Pharm. 1994;51(12):1539-54.

[16] Watkins LR, Milligan ED, Maier SF. (2003). Glial proinflammatory cytokines mediate exaggerated pain states: implications for clinical pain. *Advances in Experimental Medicine and Biology*. 521: 1-21

[17] Zhang, J.-M., & An, J. (2007). Cytokines, Inflammation and Pain. *International Anesthesiology Clinics*, *45*(2), 27–37. http://doi.org/10.1097/AIA.0b013e318034194e

[18] Werner, S., & Grose, R. (2003). Regulation of Wound Healing by Growth Factors and Cytokines. *American Physiological Society, 83*(3), 835-870.

[19] Rodgers, A. (2000). Reduction of postoperative mortality and morbidity with epidural or spinal anaesthesia: results from overview of randomised. Bmj, 321(7275), 1493-1493. 2doi:10.1136/bmj.321.7275.1493

[20] Wu, C., Hurley, R., Anderson, G., Herbert, R., Rowlingson, A., & Fleisher, L. (2004). Effect of postoperative epidural analgesia on morbidity and mortality following surgery in medicare patients. Regional Anesthesia and Pain Medicine, 29(6), 525-533. doi:10.1016/j.rapm.2004.07.002

[21] Wang J, Klein H. Red blood cell transfusion in the treatment and management of anaemia: the search for the elusive transfusion trigger. Vox Sanguinis [serial online]. January 2010;98(1):2-11. Available from: Academic Search Complete, Ipswich, MA. Accessed June 8, 2018.

[22] Seitz, K. P., Sevransky, J. E., Martin, G. S., Roback, J. D., & Murphy, D. J. (2017). Evaluation of RBC Transfusion Practice in Adult ICUs and the Effect of Restrictive Transfusion Protocols on Routine Care. *Critical Care Medicine*, *45*(2), 271–281. http://doi.org/10.1097/CCM.0000000000002077

Captions

Figure 1: Shows selection of patient cohort from MIMIC Database. After selecting those who underwent CABG procedure and excluding those with no pain measurements; 844 patients were extubated within 24 hours following surgery and included in the cohort.

Figure 2: Three plots demonstrating the bivariate relationship between the outcomes of interest and mean pain. Plot A shows decreased length of stays with increased mean pain levels. Plot B and Plot C show that, on average, those who expired at 30 days and 1 year marks experienced lower in hospital pain levels than those who did not expire.

Table 1: Shows the distribution of the outcomes and covariates in the patient cohort. Abbreviations: OASIS, Oxford Acute Severity of Illness Score; e_score, elixhauser index. OASIS score ranges from 0 to 75, with higher scores indicating more severe disease. Elixhauser index ranges from 0 to 9, with higher scores indicating a greater number of comorbid conditions.

Table 2: Shows results from main analysis and the two sensitivity analyses.
*, **, *** denotes significance at the 90%, 95%, and 99% level, respectively.

Table 3: Putative causes of truly faulty data

Table 4: Criteria to establish possible validity of counterintuitive data

Table 1: Cohort Characteristics

		No Pain	Mild	Moderate	Severe	p
n		68	419	336	21	
Age (mean (sd))		71.50 (10.61)	67.75 (10.54)	64.98 (9.73)	65.13 (12.85)	<0.001
Gender = male		45 (66.2)	333 (79.5)	282 (83.9)	14 (66.7)	0.003
oasis (mean (sd))		31.96 (7.25)	30.32 (6.47)	31.44 (6.35)	30.57 (6.20)	0.056
e_score (%)						<0.001
	0	4 (5.9)	96 (22.9)	87 (25.9)	7 (33.3)	
	1	12 (17.6)	116 (27.7)	97 (28.9)	4 (19.0)	
	2	12 (17.6)	81 (19.3)	79 (23.5)	4 (19.0)	
	3	10 (14.7)	61 (14.6)	46 (13.7)	3 (14.3)	
	4	12 (17.6)	29 (6.9)	16 (4.8)	1 (4.8)	
	5	6 (8.8)	19 (4.5)	8 (2.4)	2 (9.5)	
	6	7 (10.3)	8 (1.9)	2 (0.6)	0 (0.0)	
	7	2 (2.9)	4 (1.0)	1 (0.3)	0 (0.0)	
	8	0 (0.0)	4 (1.0)	0 (0.0)	0 (0.0)	
	9	3 (4.4)	1 (0.2)	0 (0.0)	0 (0.0)	
Mortality						
In Hospital		9 (13.2)	5 (1.2)	1 (0.3)	0 (0.0)	<0.001
30 Day		10 (14.7)	10 (2.4)	1 (0.3)	0 (0.0)	<0.001
1 Year		16 (23.5)	22 (5.3)	7 (2.1)	1 (4.8)	<0.001
Narcotics						
First 24 Hrs (sd)		4.17 (5.52)	6.24 (9.85)	9.28 (25.89)	6.38 (8.07)	0.059
Daily mean (sd)		5.23 (5.43)	8.43 (7.82)	17.09 (89.87)	8.68 (8.06)	0.162
Total		37.30 (101.39)	21.19 (70.34)	29.15 (188.08)	9.87 (8.94)	0.682

Table 2: Primary Outcome Results

Model	30 Day Mortality Odds (95% Confidence Interval)	1 Year Mortality Odds (95% Confidence Interval)	Length of Stay Estimate
Primary Analysis:			
Mean Pain	0.457***	0.710***	-0.916***
	(0.304 – 0.687)	(0.571 - 0.881)	
Median Pain	0.639***	0.856*	-0.696***
	(0.466 - 0.877)	(0.727 - 1.008)	
Max Pain	0.812***	0.887**	0.148*
	(0.693 - 0.951)	(0.790- 0.995)	
Categorical Pain	0.214***	0.450***	-2.270***
	(0.091 - 0.502)	(0.266 - 0.760)	
Sensitivity Analysis 1: Including all patients regardless of intubation length			
Mean Pain	0.592***	0.898	-0.709***
	(0.456 - 0.768)	(0.785 - 1.027)	
Categorical Pain	0.328***	0.740*	-1.706***
	(0.184 - 0.586)	(0.527 - 1.037)	
Sensitivity Analysis 2: Excluding hospital mortality patients			
Mean Pain	0.803	1.027	-0.701***
	(0.567 - 1.137)	(0.889 - 1.187)	
Categorical Pain	0.709	1.038	-1.680***
	(0.309 - 1.625)	(0.714 - 1.509)	

Table 3: Putative causes of truly faulty data

Human error	Mis-entry; misunderstanding of scale values; faulty understanding of use of data entry software; faulty interpretation of device values
Lab error	Sampling error (e.g. hemolysis); measurement error
Device error	Disconnect, interference, faulty calibration, software error; unexplained, transient aberrant values that resolve and do not recur
Systems error	Interface error, application interoperability error
Software error	Bug in software relating to data value entry; data wrongly captured, stored, and/or retrieved due to software design faults or bugs
Hardware error	Hardware issues that impact software and systems
Data analytic error	Error in analytic algorithm or process

Table 4: Criteria to establish possible validity of counterintuitive data

Viability	Is the value consistent with clinical reality? Are the values even possible ones?
Consistency	If applicable (not always the case in retrospective analysis), is the value observed consistently, such as in our pain score observations?
Continuity	What is the context of the value- does it occur as a sudden aberrant value (a 'blip'), or as one of increasingly aberrant values (a trend)?
Identity	Are the circumstances that produced the data truly identical so far as identifiable? I.e. Would the same circumstances produce the same data results in a different database, institutional, or cultural context?
Reproducibility	Is the value reproducible on repetition? while reproduction cannot be performed upon retrospective data, can the values be reproduced upon observation across different clinical databases, or in the same database over ongoing time?
Sensibility	Even if it does not meet current clinical expectations, does it make potential sense in associated clinical context?
Curiosity	Does it drive the observer to seek alternative better solutions and pose questions for further research?

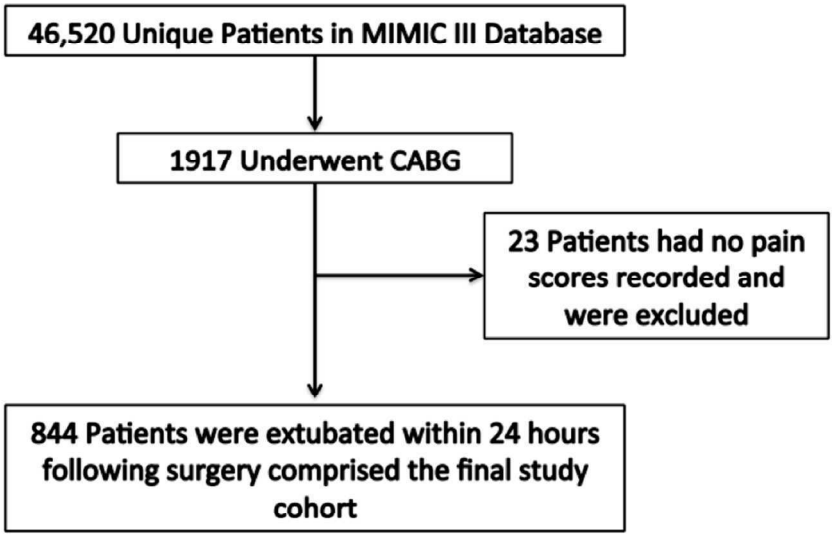
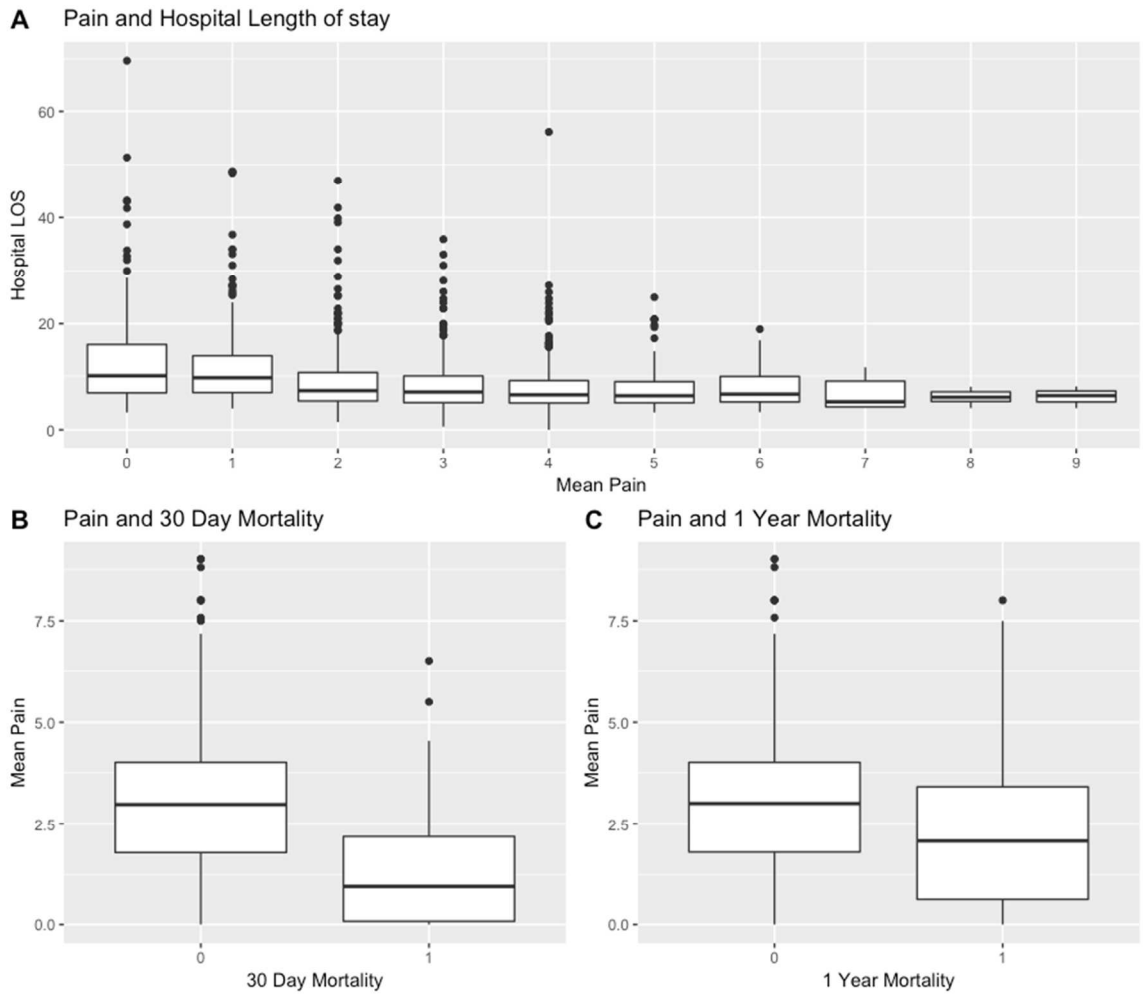


Figure 1:

Figure 2: Bivariate Exploration



STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*

Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any pre-specified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	3-4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	4
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4-5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*4-5*
Bias	9	Describe any efforts to address potential sources of bias	9
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6
		(*b*) Describe any methods used to examine subgroups and interactions	6
		(*c*) Explain how missing data were addressed	6
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	NA

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
		(*e*) Describe any sensitivity analyses	6
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	NA
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6-7, Table 1
		(b) Indicate number of participants with missing data for each variable of interest	NA
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	8, Table 2
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Table 2
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	7
Discussion			
Key results	18	Summarise key results with reference to study objectives	7
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	9
Generalisability	21	Discuss the generalisability (external validity) of the study results	9
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	13

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.
Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Counterintuitive Results From Big Data: A Case Study and Discussion

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-026447.R1
Article Type:	Research
Date Submitted by the Author:	07-Jan-2019
Complete List of Authors:	Doty, Erik; Harvard Medical School Department of Biomedical Informatics Stone, David; University of Virginia School of Medicine; Massachusetts Institute of Technology, Laboratory for Computational Physiology McCague, Ned; Massachusetts Institute of Technology, Laboratory for Computational Physiology; Kyruus, Inc Celi, Leo Anthony; Massachusetts Institute of Technology,
Primary Subject Heading:	Intensive care
Secondary Subject Heading:	Health informatics
Keywords:	Pain, Mortality, Length of Stay, Big Data, Counterintuitive results

Counterintuitive Results From Big Data: A Case Study and Discussion

Erik Doty[1], DO, MBI; David J. Stone[2,5], MD; Ned McCague[3,5], MPH; Leo Anthony Celi[4,5], MD, MS, MPH

Corresponding author: Leo Anthony Celi

Email: lceli@bidmc.harvard.edu

Phone: (617) 253-7937

1. *Harvard Medical School Department of Biomedical Informatics, Boston, MA*
2. *University of Virginia School of Medicine, Charlottesville, VA*
3. *Kyruus, Inc., Boston, MA*
4. *Beth Israel Deaconess Medical Center, Boston, MA*
5. *MIT Laboratory for Computational Physiology, Cambridge, MA*

Counterintuitive Results from Big Data: A Case Study and Discussion

E. Doty, D.J. Stone, N. McCague, L.A. Celi

Keywords: Pain, mortality, length of stay

Abstract

Objective: Explore the issue of counterintuitive data via analysis of a representative case in which the data obtained was unexpected and apparently inconsistent with current knowledge. We then discuss the general issue of counterintuitive data while developing a framework for approaching such findings.

Design: The case was a retrospective analysis of a cohort of Coronary Artery Bypass Graft (CABG) patients. Regression was used to examine the association between perceived pain in the ICU and selected outcomes.

Setting: MIMIC-III, a publicly available, deidentified critical care patient database.

Participants: 844 adult patients from the database who underwent CABG surgery and were extubated within 24 hours after ICU admission.

Outcomes: 30 Day mortality, 1-year mortality, and hospital length of stay (LOS).

Results: Increased pain levels were found to be significantly associated with reduced mortality at 30 days and 1-year, and shorter hospital LOS. A one-point increase in mean pain level was found to be associated with a reduction in the odds of 30-day and 1-year mortality by a factor of 0.457 (95%CI 0.304-0.687, $p< 0.01$) and 0.710 (95%CI 0.571 - 0.881, $p< 0.01$) respectively, and a 0.916 (95%CI (-1.159, -0.673), $p < 0.01$) day decrease in hospital LOS.

Conclusion: The finding of an association between increased pain and improved outcomes was unexpected and clinically counterintuitive. In an increasingly digitized age of medical big data, such results are likely to become more common. The reliability of such counterintuitive results must be carefully examined: We suggest several issues to consider in this analytic process. If the data is determined to be valid, consideration must then be made towards alternative explanations for the counterintuitive results observed. Such results may in fact indicate that current clinical knowledge is incomplete or not have been firmly based on data, and function to inspire further research into the factors involved.

Strengths and limitations of this study

- Large sample size with minimal covariate data missing.
- Multiple regression models with multiple sensitivity analyses.
- High internal validity shown by use of falsification hypothesis testing.
- Lack of oral analgesic data.

- Recognizing that correlation does not equal causation and further work is needed to confirm case results.

Introduction

What do we mean by counterintuitive data? It is data that presents unexpected results that may clash with common sense or what has been previously published and accepted by the medical community. In practice, clinicians have long dealt with such results in individual bits but have had the vast advantage of being able to examine the concurrent state of the patient and react in real time by repeating a lab test or tracking ongoing monitor data. These responses function to identify the prior result as a non-repeatable error, or as a genuine anomaly. However, this approach is not applicable to the context of retrospective data analysis. Furthermore, the counterintuitive data revealed in such analyses is likely to be more involved than a single aberrant lab or vital sign value. In today's data driven healthcare system, retrospective data analyses are becoming more and more common. We can therefore logically expect to encounter a greater incidence and variety of counterintuitive values and results that are impossible to confirm by repetition, difficult to confirm or deny by context, but still require interpretation.

The question then becomes how best to approach such results? Are they incorrect simply because they weren't what was expected? And was the expectation itself based on subjective assumptions or objective conclusions? When our prior expectations are not met, are we dealing with truly faulty data, or do our expectations need to be reset by what are reliable, but counterintuitive, results. For example, we have learned that intensive care practices common in the past such as large tidal volume ventilation, the use of pulmonary artery catheters, and the use of lidocaine infusions in myocardial infarction led to no benefit or injury.[1-3] Were these unexpected negative outcomes initially missed because outcomes data was not being carefully analyzed, or perhaps ignored or interpreted as counterintuitive to the level of unbelievability? How can the situation be dissected retrospectively so that counterintuitive data can be identified as truly spurious versus simply not being consistent with our prior experience which may itself be faulty and require data driven correction?

In this paper, we explore a case in which the results contradicted previous reports and our

clinical expectations. Using the Medical Information Mart for Intensive Care-III (MIMIC-III), a critical care database that was developed and maintained by the Laboratory for Computational Physiology at the Massachusetts Institute of Technology[4], we retrospectively selected a cohort of patients that underwent a coronary artery bypass graft (CABG) procedure and evaluated the effect of perceived pain on mortality and hospital length of stay (LOS). Our initial hypothesis was that increased levels of perceived pain would correlate with worse patient outcomes such as increased hospital length of stay. This would be in line with the current literature that suggest optimal pain control leads to increased mobility, earlier ambulation, and improved outcomes. [5-7] Contrary to the literature, we found that higher levels of pain were associated with reduced mortality and reduced LOS. We then discuss potential causes of these results and the general issue of dealing with counterintuitive results in retrospective data analyses.

Case

Population

We selected patients from the MIMIC database who met all of the following inclusion criteria and none of the exclusion criteria. Inclusion criteria included: (1) Adult > 18 years old, (2) who underwent CABG surgery, and (3) was extubated within 24 hours after arrival to the ICU. Exclusion criteria were: (1) Non-CABG surgical procedure, and (2) missing data on confounding variables. Patients were identified using Current Procedural Terminology (CPT) codes: The following CPT codes corresponded to the CABG procedure: 33510 to 33516 for venous grafting only for coronary artery bypass, and 33533 to 33548 for arterial grafting for coronary bypass. The final study cohort contained 844 patients (***Figure 1***).

The MIMC-III database included 1,917 patients who underwent CABG, with 844 meeting the study criteria. CABG was chosen for the investigation as it is a common procedure with the majority of patients having no or few post-operative complications and relatively predictable recoveries.[5] Due to the nature of the surgical procedure which requires sternal spreading for exposure, there is an expected high analgesic burden immediately after surgery.

Outcomes

The primary outcome assessed was mortality at 30 days. Secondary outcomes were mortality at 1 year and hospital LOS. In the MIMIC database, mortality data for patients who die after hospital

discharge is derived from the social security death registry.[4]

Exposures

The exposures of interest were pain levels reported by the patient immediately and in the subsequent interval after ICU extubation. Pain levels on a scale of 0-10 were regularly self-reported by patients to ICU nurses and recorded in the database, generating a continuum of measurements for each patient. The mean, median, and maximum pain levels were used for separate analyses. Concomitant measurements of heart rates, respiratory rates, and systolic blood pressures were also compared against their simultaneously recorded pain measurement.

Intravenous (IV) opiate administration was extracted from the database. MIMIC contained data for the following medications: Morphine, fentanyl, hydromorphone, and meperidine. The was no data in MIMIC corresponding to the administration of oral analgesics.

Nausea and delirium were also tested against our outcomes for use in falsification hypothesis testing. The presence of nausea was derived from the nursing notes stored in the database. A positive nausea exposure was defined as the mention “nausea” or “nauseous” in the nursing note with no negative descriptor, such as “not nauseous” or “denies nausea”, attached. Delirium was similarly assessed by looking for mention of “delirium”, “delirious”, or “confusion”. Additionally, delirium exposure was considered positive if patients had a positive nursing delirium assessment.

Covariates

Several variables found to be linked to worse patient outcomes in previous studies were included to control for confounding in the regression models: demographic factors, comorbid conditions, and illness severity score on admission to the ICU.[8,9] Comorbid burden was represented by the Elixhauser index which is determined by the aggregate presence or absence of 30 different comorbid conditions as detected by ICD-9 codes.[10] These conditions include but are not limited to cardiovascular disorders such as hypertension, congestive heart failure, coronary artery disease, and peripheral vascular disease; pulmonary disorders such as chronic obstructive pulmonary disease; endocrine disorders such as diabetes and hypothyroid; obesity; drug and alcohol use disorders; renal disease; liver disease. Illness severity was captured using the Oxford Acute Severity of Illness Score (OASIS), which is calculated on admission to the ICU and takes

into account age, heart rate, Glasgow coma scale, mean arterial pressure, temperature, respiratory rate, ventilatory status, urine output, pre-ICU in-hospital LOS, and whether or not the patient underwent elective surgery. Studies have shown OASIS is comparable to other illness severity ratings in predicting outcomes such as mortality and length of stay.[11]

Patient and Public Involvement

This research was done without patient or public involvement. They were not invited to contribute to the development of our methodology, our outcomes, nor the writing of our paper.

Statistical Analysis

Analysis was carried out using R version 3.4.0 and SAS 9.4. Binomial logistic regression models were fitted using maximum likelihood estimation to compare the pain measures with 30-day and 1-year mortality. Linear regression was used to model the relationship between mean pain scores and hospital LOS. Age, gender (male reference), Elixhauser index, and OASIS score were included in the models to account for potential confounders. In a separate regression, mean pain levels were categorized into four ordinal groups of no pain (0/10), mild pain (1-3), moderate pain (3-6), and severe pain (7-10) in accordance with the NIH Pain Consortium.[12]

ANOVA was used to determine if there was a significant variation in heart rate, respiratory rate, and/or systolic blood pressure, when compared to the concurrent pain assessment.

IV analgesia medications were converted to their morphine equivalents based on the National Pharmaceutical Counsel's guidelines.[13] The IV analgesia was subdivided into total dose in the first 24 hours, mean dose per ICU course day, and total dose during ICU course. ANOVA models were used to determine if there were any significant variation in administration of IV analgesics among the four categorized pain groups.

Two sensitivity analyses were performed to assess the robustness of the observed effects. The first included the same statistical tests in all postoperative CABG patients regardless of duration of intubation. The second sensitivity analysis excluded patients who died in the hospital.

To add validity to the potential observed associations, falsification hypothesis testing using Prasad and Jena's methodology was employed. A distinct and highly unlikely hypothesis is

tested against the outcomes of interest.[14] In our case, we used nausea, a symptom with no known effect on clinical outcomes, and tested it against mortality and length of stay. We also tested delirium, a symptom associated with poorer patient outcomes, against the outcome measures.

Results

The database included 844 patients who underwent a CABG procedure and were extubated within 24 hours. There were 68 patients who on average reported no pain during their ICU stay after extubation, 419 with mild pain, 336 with moderate pain, and 21 with severe pain. The mean frequency of pain measurements was 19.8 measurements per patient. The distribution of patient characteristics, including age, gender, illness acuity on ICU admission (OASIS), and comorbidity index is reported in ***Table 1***. There was no significant difference noted in the frequency in which pain was assessed in those who experienced lower pain levels when compared to those who experienced increased pain levels. The number of comorbidities ranged from 0 to 9. Bivariate analysis showed increasing OASIS was significantly associated with increased mortality and increased LOS ($p < 0.05$). No significant differences were found in the amount of IV analgesia administered among the pain subgroups.

		No Pain	Mild	Moderate	Severe	p
n		68	419	336	21	
Age (mean (sd))		71.50 (10.61)	67.75 (10.54)	64.98 (9.73)	65.13 (12.85)	<0.001
Gender = male (%)		45 (66.2)	333 (79.5)	282 (83.9)	14 (66.7)	0.003
oasis (mean (sd))		31.96 (7.25)	30.32 (6.47)	31.44 (6.35)	30.57 (6.20)	0.056
e_score (%)						<0.001
	0	4 (5.9)	96 (22.9)	87 (25.9)	7 (33.3)	
	1	12 (17.6)	116 (27.7)	97 (28.9)	4 (19.0)	
	2	12 (17.6)	81 (19.3)	79 (23.5)	4 (19.0)	
	3	10 (14.7)	61 (14.6)	46 (13.7)	3 (14.3)	
	4	12 (17.6)	29 (6.9)	16 (4.8)	1 (4.8)	
	5	6 (8.8)	19 (4.5)	8 (2.4)	2 (9.5)	
	6	7 (10.3)	8 (1.9)	2 (0.6)	0 (0.0)	
	7	2 (2.9)	4 (1.0)	1 (0.3)	0 (0.0)	
	8	0 (0.0)	4 (1.0)	0 (0.0)	0 (0.0)	
	9	3 (4.4)	1 (0.2)	0 (0.0)	0 (0.0)	
Mortality						

In Hospital (%)	9 (13.2)	5 (1.2)	1 (0.3)	0 (0.0)	<0.001
30 Day (%)	10 (14.7)	10 (2.4)	1 (0.3)	0 (0.0)	<0.001
1 Year (%)	16 (23.5)	22 (5.3)	7 (2.1)	1 (4.8)	<0.001
Narcotics					
First 24 Hrs (sd)	4.17 (5.52)	6.24 (9.85)	9.28 (25.89)	6.38 (8.07)	0.059
Daily mean (sd)	5.23 (5.43)	8.43 (7.82)	17.09 (89.87)	8.68 (8.06)	0.162
Total Narcotics (sd)	37.30 (101.39)	21.19 (70.34)	29.15 (188.08)	9.87 (8.94)	0.682

Table 1: Shows the distribution of the outcomes and covariates in the patient cohort. Abbreviations: OASIS, Oxford Acute Severity of Illness Score; e_score, elixhauser index. OASIS score ranges from 0 to 75, with higher scores indicating more severe disease. Elixhauser index ranges from 0 to 9, with higher scores indicating a greater number of comorbid conditions.

Bivariate analysis (***Figure 2***) shows a correlation between increasing pain levels and improved outcomes among these patients who had no intra-operative complications and were extubated within 24 hours of arrival in the ICU. Higher pain levels for this specific cohort of patients who were fast-tracked after CABG were found to be associated with decreased hospital LOS. Those who experienced lower levels of pain in the ICU were more likely to be dead at 30 days and 1 year.

Multivariate regression analysis was performed to adjust for confounding. Four different models using mean, median, and maximum pain scores, and pain categories were tested against the clinical outcomes with the results displayed in ***Table 2***. The logistic regression models consistently showed that increasing pain was associated with reduced odds of death at 30 days and 1 year after adjustment for illness severity and co-morbid conditions. All the linear models demonstrated that increasing pain levels were also associated with decreased hospital LOS, except for the model that looked at the maximum pain score, which showed an opposite effect. R-Squared values for the linear regression models varied between 0.25 and 0.3 for all the models. Complete statistical data from all regression models can be found in the **online supplemental materials file.**

Model	30 Day Mortality Odds (95% Confidence Interval)	1 Year Mortality Odds (95% Confidence Interval)	Length of Stay Estimate (95% Confidence Interval)

Primary Analysis:			
Mean Pain	0.457*** (0.304 – 0.687)	0.710*** (0.571 - 0.881)	-0.916*** (-1.159, -0.673)
Median Pain	0.639*** (0.466 - 0.877)	0.856* (0.727 - 1.008)	-0.696*** (-0.886, -0.506)
Max Pain	0.812*** (0.693 - 0.951)	0.887** (0.790- 0.995)	0.148* (-0.02, 0.32)
Categorical Pain	0.214*** (0.091 - 0.502)	0.450*** (0.266 - 0.760)	-2.270*** (-2.903, 1.637)
Sensitivity Analysis 1: Including all patients regardless of intubation lengths			
Mean Pain	0.592*** (0.456 - 0.768)	0.898 (0.785 - 1.027)	-0.709*** (-0.866, -0.552)
Categorical Pain	0.328*** (0.184 - 0.586)	0.740* (0.527 - 1.037)	-1.706*** (-2.110, -1.302)
Sensitivity Analysis 2: Excluding hospital mortality patients			
Mean Pain	0.803 (0.567 - 1.137)	1.027 (0.889 - 1.187)	-0.701*** (-0.858, -0.544)
Categorical Pain	0.709 (0.309 - 1.625)	1.038 (0.714 - 1.509)	-1.680*** (-2.082, -1.278)

Table 2: Shows results from main analysis and the two sensitivity analyses.
*, **, *** denotes significance at the 90%, 95%, and 99% level, respectively.

No significant variations were noted in heart rate, respiratory rate, or blood pressure with increasing pain levels.

Sensitivity analysis was employed to examine all patients regardless of duration of intubation, expanding the sample size to 1889 patients. The results were similar for 30-day mortality and hospital LOS as regards effect size and statistical significance; however, the results were not statistically significant for 1-year mortality (***Table 2***). An additional sensitivity analysis excluded patients who died in the hospital- these results were consistent with the prior models and were statistically significant for hospital LOS, but not for mortality (***Table 2***).

As expected, the presence of nausea was not found to be associated with any impact on outcomes in the study cohort. As also would be expected, patients who had delirium had worse 30-day and

1-year mortality and longer hospital LOS. This helps support that the above observations of pain levels and its effect on the outcomes are less likely due to chance.

Discussion

Case

We will first discuss our unexpected results, and then discuss the general issue of counterintuitive data. Our results that increasing levels of patient-reported pain severity post-CABG surgery are associated with better clinical outcomes were not consistent with our initial hypothesis that better outcomes would correlate with better pain control as per the reported literature. In fact, prior studies have found increased levels of pain in the hospital to be associated with increased mortality. [15]

The difference in the study cohort between our study and others may explain some of the discordance. Our initial analysis was limited to "fast-tracked" patients who did not have intra-operative complications and were extubated early in their ICU course. These patients made up 44% of the database patients. Studies that have reported worse clinical outcomes associated with post-operative pain did not select for a relatively healthy sub-cohort of patients. Why would patients with higher levels of pain have better outcomes? It is well documented that an increased inflammatory reaction is associated with increased pain. Pro-inflammatory cytokines such as IL-1β, IL-6, and TNF-α have been directly implicated in the physiology of pain.[16,17] These cytokines have also been found to be directly involved in wound healing through the stimulation of processes such as keratinocyte and fibroblast proliferation, and synthesis and breakdown of extracellular matrix proteins.[18] We speculate that those patients who demonstrated better outcomes mounted a more robust inflammatory response leading to more pain, but also to increased healing ability.

Another possibility is that higher perceived pain levels represent a proxy for a generally better state of health, including superior physiological function of the cardiovascular, respiratory, renal, and hepatic systems. In tandem, these systems act to metabolize and eliminate anesthetic and analgesic drugs so that the net pharmacokinetic result would likely be increased susceptibility to pain due to less administered agent remaining at active sites. Furthermore, patients with better cardiovascular function would likely have better cerebral perfusion with improved central

neurological function, and thereby have a pharmacodynamic reason for perceiving more pain. And patients who are generally in better overall condition would be expected to manifest better outcomes. These thoughts are admittedly speculative and additional research is needed to explore these possibilities.

It is important to point out that the goal of clinicians should not be in any way to maximize pain to optimize outcomes. Conventional approaches that aim to control pain adequately should be employed. Our observation is just that - an observation of an association and conjectures of possible linking mechanisms but is not intended in any way to drive pain management policy in the direction of tolerating undertreated pain.

We performed sensitivity analyses, one including all patients regardless of post-operative ventilation duration, and another excluding patients who died during hospitalization, and reached similar conclusions. When excluding in-hospital deaths, we discovered the 30-day mortality rate had a similar odds ratio but was no longer statistically significant. This is most likely due to the low mortality rate after hospital discharge following CABG, making it difficult to detect a statistically significant effect.

We believe that researcher bias is a non-issue as these findings were not expected, but rather, the opposite. Sampling bias was also minimal. Our inclusion criteria were predefined prior to database sampling and only 28 patients needed to be excluded due to missing data. We performed multiple sensitivity analyses to determine if those that were excluded would have influenced our results. However, the study has several limitations inherent in any retrospective data analysis. We acknowledge that correlation does not equal causation and further research is needed to determine the underlying physiologic mechanism for the results seen. Due to the self-reported nature of the pain scores, reporting bias is a concern. Some patients may have over-reported and others under-reported their pain. We also recognize that analgesic administration is a confounder. While we were unable to completely control for this due to lack of information regarding oral analgesics in the database However, with respect to intravenous analgesics, we attempted to limit this potential confounder by excluding those with prolonged intubations who would inherently have received and required greater doses of sedatives and analgesics. We also compared the amount of narcotics that patients were receiving and did not observe any significant differences among the various pain groups. Despite measures taken to guarantee

internal validity, we anticipate appropriate skepticism with regard to generalizability of the findings. This, of course, is of genuine concern given the current state-of-affairs where clinicians are already inundated with conflicting studies of questionable quality. We therefore invite other investigators to replicate (and expand) our analysis in other databases.

Counterintuitive Results and other examples

As noted, our findings were contrary to clinical expectations and to most published works which associate increased pain with worse outcomes.[15,19-20] Encountering counterintuitive results is not unique to retrospective data analysis. Clinicians encounter unexpected, possibly aberrant, values in situations such as the evaluation of laboratory and monitor data. When a possibly spurious lab result is obtained, the usual response is to repeat the test. When the second test comes back with a more acceptable value, we generally then ignore the unexpected value. But what if the repeat value is also aberrant? Do we repeat it again, or do we begin to believe that the value is 'real' and start to formulate a response to a clinical problem? In this case, it is the *consistency* and *reproducibility* of the counterintuitive value that drives its possible validity. The details of this process are determined by the overall clinical risks involved. The consistency we found in the pain score values drove us to consider the possibility that the values were 'real' even though they were counterintuitive in terms of our expectations.

Another issue in evaluating to counterintuitive values is whether they are *possible.* Impossible values would include a potassium of 64.5, one incompatible with life. But a potassium of 7.3 is a possible value. The pain values associated with better outcomes were unexpected, but not so high that they were impossible in the observed context.

One question that would arise with a potassium of 7.3 would be that of continuity- did the value occur suddenly or gradually in a stream of normal values? Were surrounding values similarly abnormal? In the context of persistently abnormal values, e.g. untreated uremia, a normal value would be counterintuitive. So that while most counterintuitive values will tend to be out of the 'normal range', they will not necessarily be so. In the context of increasing values, it might simply be the first one that was not only out of the normal range, but that crossed the line into a critical range,

The fundamental question is whether counterintuitive results are actually false results, or does the problem lie in our perception of what should be. **Table 3** displays a categorization of error types that could result in faulty data. We are not able to attribute the counterintuitive data we observed to any of these factors, however.

Human error	Mis-entry; misunderstanding of scale values; faulty understanding of use of data entry software; faulty interpretation of device values
Lab error	Sampling error (e.g. hemolysis); measurement error
Device error	Disconnect, interference, faulty calibration, software error; unexplained, transient aberrant values that resolve and do not recur
Systems error	Interface error, application interoperability error
Software error	Bug in software relating to data value entry; data wrongly captured, stored, and/or retrieved due to software design faults or bugs
Hardware error	Hardware issues that impact software and systems
Data analytic error	Error in analytic algorithm or process

Table 3: Putative causes of truly faulty data

How can counterintuitive results be approached in secondary data analyses? **Table 4** displays characteristics that may distinguish reliable (but counterintuitive) from truly faulty data. With consideration of these factors, the first investigative step is to retrace the process and workflow involved in data entry so far as possible. Our data was obtained at the institution of several of the authors where nurses are trained to assess pain on a standard scale from 0 to 10. There are several potential faults to this method. The nursing staff could neglect to regularly assess pain or neglect to enter the information into the medical record generating the database. While this may alter a few data points, it is unlikely to systematically affect all data unless there was an obvious glaring institutional issue affecting every nurse and every data entry.

Viability	Is the value consistent with clinical reality? Are the values even possible ones?
Consistency	If applicable (not always the case in retrospective analysis), is the value observed consistently, such as in our pain score observations?
Continuity	What is the context of the value- does it occur as a sudden aberrant value (a 'blip'), or as one of increasingly aberrant values (a trend)?
Identity	Are the circumstances that produced the data truly identical so far as identifiable? I.e. Would the same circumstances produce the same data results in a different database, institutional, or cultural context?
Reproducibility	Is the value reproducible on repetition? while reproduction cannot be performed upon retrospective data, can the values be reproduced upon observation across different clinical databases, or in the same database over ongoing time?
Sensibility	Even if it does not meet current clinical expectations, does it make potential sense in associated clinical context?
Curiosity	Does it drive the observer to seek alternative better solutions and pose questions for further research?

Table 4: Criteria to establish possible validity of counterintuitive data

After determining that the data source is valid, additional statistical tests can be run on the patient cohort. Tests such as the falsification hypothesis testing we utilized, add validity to the results as they show that the cohort follows other generally known principles. In our study, falsification analysis by both neutral (nausea) and positive (delirium) factors provided support for our findings.

Concurrent contextual data can also help to confirm the veracity of data- for example, one could examine ECGs if hyperkalemia was being analyzed. We examined concomitant vital signs

during the time of pain measurements. We expected to observe significant increases with higher pain levels, but did not: With the combination of analgesics, residual anesthetics, and the concurrent use of drugs that directly affect vital signs such as beta-blockers, the lack of correlation is probably not surprising. In fact, we learned that in this setting, it appears to be inadvisable to use vital sign changes as a proxy for the presence of unvoiced pain. Finally, one can attempt to physiologically explain the disparity between the observed and expected results as we did above for the case of post- CABG pain.

The use of lower thresholds for blood transfusions in the ICU is an example of a counterintuitive finding. ICU target hemoglobin levels were historically set at greater than 10 g/dL, theoretically to ensure adequate oxygen delivery.[21] This led to increased transmission of blood borne diseases, unnecessary healthcare expenditures, and actually worse outcomes.[22] Later research showed that this rule was not necessary for most patients, but only for selected patients such as those with acute coronary syndrome actively experiencing chest pain. The initially counterintuitive findings that lower hemoglobin levels were not only acceptable but preferable in most cases, served as research triggers to more fully elucidate optimal clinical practice. Our case may serve as an analogous research trigger in terms of optimally managing postoperative pain. Outcomes such as mortality and LOS are complex phenomena driven by many factors- to observe a clear and robust statistical effect such as we did is strongly suggestive that something 'real' is occurring even if the data were counterintuitive.

The final step when dealing with counterintuitive data is to look for additional evidence that confirms the reliability of the results (perhaps this could be termed 'confirmatory metadata'). With respect to our CABG case, the analysis should be rerun on additional databases and in different settings. Just as clinicians continued to manage intensive care unit anemia as they always had until more definitive results were reported, our results should not impact the analgesic care of patients at this point. However, we hope that we have raised the issue in the appropriate minds that outcomes may benefit from approaches slightly different from usual. After all, one can easily eliminate all pain from postoperative patients but they would have to remain sedated and ventilated for an indefinite period of time to do so. And after they are extubated, pain management should not be so aggressive that it leads to apnea and respiratory arrest. In other words, there may be a detectable level of tolerable pain that leads patients to their

best outcomes, and no honest clinician will guarantee a patient that they will have no pain at all after a procedure like a sternal-disrupting CABG.

Conclusion

Contrary to our expectations, we observed, in a retrospective analysis of electronic health records, that post-CABG fast-track patients with higher pain scores had better outcomes. The increasing use of EHRs for secondary analysis will likely lead to an increasing incidence of such apparently counterintuitive results. While the first step in this situation is to attempt to confirm the reliability of both the analytic process and the data itself, such findings that prove to be robust may lead to further ideas and subsequent research that drive future clinical care. On the other hand, clinicians must be careful in terms of modifying their practices until the implications of such counterintuitive (or any) data have been thoroughly vetted and confirmed in diverse database contexts and via the peer review process.

Declarations

Availability of Data Materials

The datasets generated for the current study were derived from the MIMIC-III Database available at https://mimic.physionet.org/. The data subsets and statistical code used in this project can be found at https://github.com/ErikWDoty/PainProject.

Consent for Publication

Not applicable

Competing Interests

The authors declare they have no completing interests.

Funding

There was no financial support for this project.

Author Contributions

ED was responsible for the data extraction, the initial statistical analysis, and writing and editing the manuscript. NM was involved in validating the statistical models and participated in editing

the manuscript. DS was responsible for assisting with background information and editing the manuscript. LC was the project supervisor, responsible for project conception and manuscript editing.

Acknowledgements

We would like to thank J. Michael Jaeger, MD, PhD of the University of Virginia School of Medicine, for his assistance with background information.

References

[1] Gajic O, Dara SI, Mendez JL, et al. Ventilator-associated lung injury in patients without acute lung injury at the onset of mechanical ventilation. Critical Care Medicine. 2004;32:1817-1824.

[2] A.F Connors Jr, T Speroff, N.V Dawson, et al. The effectiveness of right heart catheterization in the initial care of critically ill patients (SUPPORT Investigators) JAMA, 276 (1996), pp. 889-89

[3] Hine LK, Laird N, Hewitt P, Chalmers TC. Meta-analytic Evidence Against Prophylactic Use of Lidocaine in Acute Myocardial Infarction. Arch Intern Med. 1989;149(12):2694–2698. doi:10.1001/archinte.1989.00390120056011

[4] MIMIC-III, a freely accessible critical care database. Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. Scientific Data (2016). DOI: 10.1038/sdata.2016.35. Available from: http://www.nature.com/articles/sdata201635

[5] Sutherland, J., Liu, G., Crump, T., Bair, M., & Karimuddin, A. (2018). Relationship between preoperative patient-reported outcomes and hospital length of stay: A prospective cohort study of general surgery patients in Vancouver, Canada. *Journal of Health Services Research & Policy*. doi:10.1177/1355819618791634

[6] Bykowski, M. R., Sivak, W., Garland, C., Cladis, F. P., Goldstein, J. A., & Losee, J. E. (2018). A Multimodal Preemptive Analgesic Protocol for Alveolar Bone Graft Surgery: Decreased Pain, Hospital Stay, and Health Care Costs. *The Cleft Palate-Craniofacial Journal*. doi:10.1177/1055665618791943

[7] Ljungqvist O, Scott M, Fearon KC. (2017) Enhanced Recovery After Surgery: A

Review. *JAMA Surg*;152(3):292–298. doi:10.1001/jamasurg.2016.4952

[8] Hawkes, A. L., Nowak, M., Bidstrup, B., & Speare, R. (2006). Outcomes of coronary artery bypass graft surgery. Vascular Health and Risk Management, 2(4), 477–484.Hawkes, A. L., Nowak, M., Bidstrup, B., & Speare, R. (2006). Outcomes of coronary artery bypass graft surgery. *Vascular Health and Risk Management*, *2*(4), 477–484.

[9] Shroyer, A. W., Coombs, L. P., Peterson, E. D., Eiken, M. C., Delong, E. R., Chen, A., Edwards, F. H. (2003). The society of thoracic surgeons: 30-day operative mortality and morbidity risk models. The Annals of Thoracic Surgery, 75(6), 1856-1865. doi:10.1016/s0003-4975(03)00179-6

[10] Elixhauser, A., Steiner, C., Harris, D., & Coffey, R. (1998). Comorbidity Measures for Use with Administrative Data. Medical Care, 36(1), 8-27. Retrieved from http://www.jstor.org/stable/3766985

[11] Johnson, A. E., Kramer, A. A., & Clifford, G. D. (2013). A New Severity of Illness Scale Using a Subset of Acute Physiology and Chronic Health Evaluation Data Elements Shows Comparable Predictive Accuracy*. Critical Care Medicine, 41(7), 1711-1718. doi:10.1097/ccm.0b013e31828a24fe

[12] Pain Intensity Instruments. (2003, July). Retrieved April 9, 2017, from https://painconsortium.nih.gov/pain_scales/NumericRatingScale.pdf

[13] *Pain: Current understanding of assessment, management, and treatments*. (2001). Reston, VA: National Pharmaceutical Council.

[14] Prasad, V., & Jena, A. B. (2013). Prespecified Falsification End Points. *Jama, 309*(3), 241. doi:10.1001/jama.2012.96867

[15] Lewis KS, Whipple JK, Michael KA, Quebbeman EJ. Effect of analgesic treatment on

the physiological consequences of acute pain. Am J Hosp Pharm. 1994;51(12):1539-54.

[16] Watkins LR, Milligan ED, Maier SF. (2003). Glial proinflammatory cytokines mediate exaggerated pain states: implications for clinical pain. *Advances in Experimental Medicine and Biology*. 521: 1-21

[17] Zhang, J.-M., & An, J. (2007). Cytokines, Inflammation and Pain. *International Anesthesiology Clinics*, *45*(2), 27–37. http://doi.org/10.1097/AIA.0b013e318034194e

[18] Werner, S., & Grose, R. (2003). Regulation of Wound Healing by Growth Factors and Cytokines. *American Physiological Society, 83*(3), 835-870.

[19] Rodgers, A. (2000). Reduction of postoperative mortality and morbidity with epidural or spinal anaesthesia: results from overview of randomised. Bmj, 321(7275), 1493-1493. 2doi:10.1136/bmj.321.7275.1493

[20] Wu, C., Hurley, R., Anderson, G., Herbert, R., Rowlingson, A., & Fleisher, L. (2004). Effect of postoperative epidural analgesia on morbidity and mortality following surgery in medicare patients. Regional Anesthesia and Pain Medicine, 29(6), 525-533. doi:10.1016/j.rapm.2004.07.002

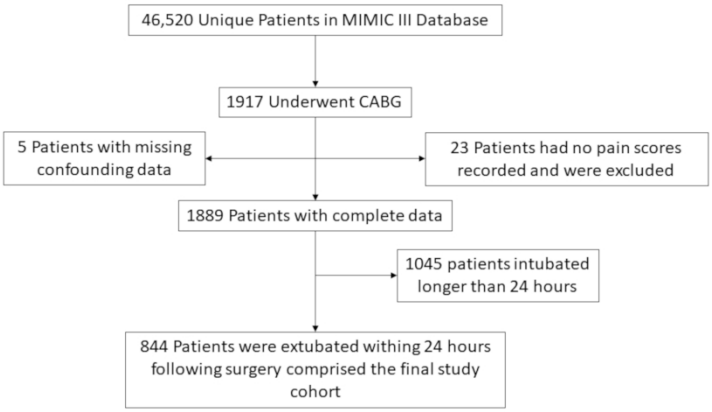
[21] Wang J, Klein H. Red blood cell transfusion in the treatment and management of anaemia: the search for the elusive transfusion trigger. Vox Sanguinis [serial online]. January 2010;98(1):2-11. Available from: Academic Search Complete, Ipswich, MA. Accessed June 8, 2018.

[22] Seitz, K. P., Sevransky, J. E., Martin, G. S., Roback, J. D., & Murphy, D. J. (2017). Evaluation of RBC Transfusion Practice in Adult ICUs and the Effect of Restrictive Transfusion Protocols on Routine Care. *Critical Care Medicine*, *45*(2), 271–281. http://doi.org/10.1097/CCM.0000000000002077

Captions

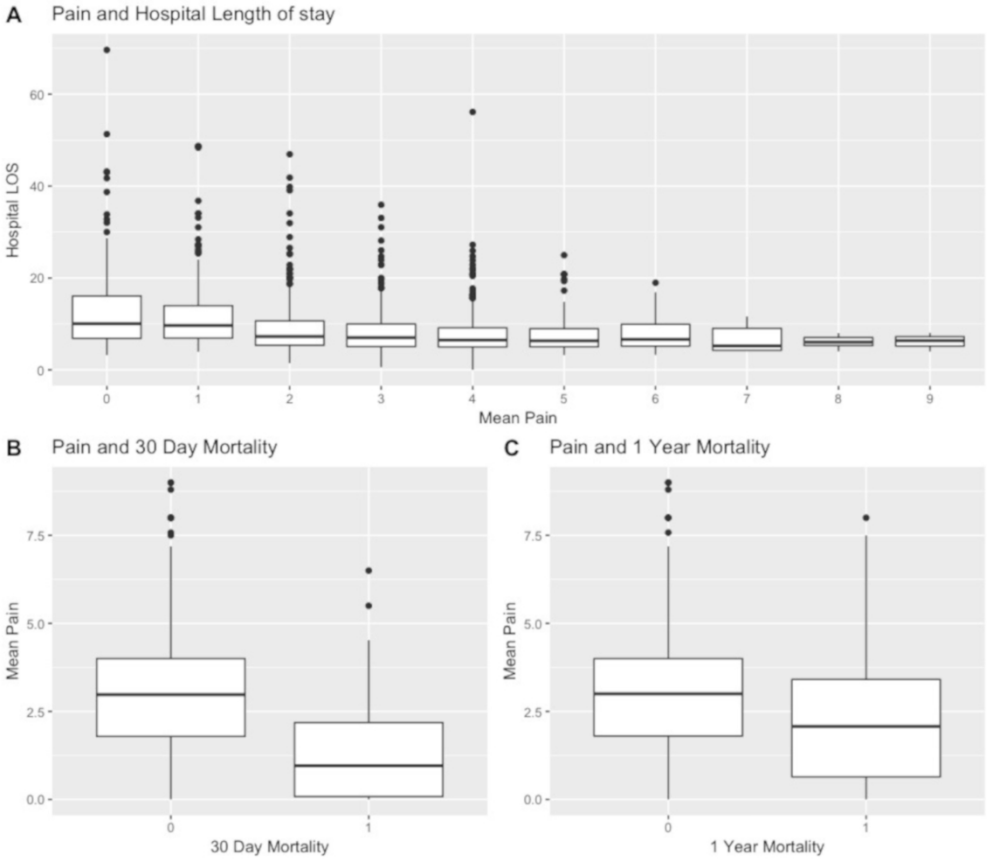
Figure 1: Shows selection of patient cohort from MIMIC Database. After selecting those who underwent CABG procedure and excluding those with no pain measurements; 844 patients were extubated within 24 hours following surgery and included in the cohort.

Figure 2: Three plots demonstrating the bivariate relationship between the outcomes of interest and mean pain. Plot A shows decreased length of stays with increased mean pain levels. Plot B and Plot C show that, on average, those who expired at 30 days and 1 year marks experienced lower in hospital pain levels than those who did not expire.



Shows selection of patient cohort from MIMIC Database. After selecting those who underwent CABG procedure and excluding those with no pain measurements; 844 patients were extubated within 24 hours following surgery and included in the cohort.

177x99mm (600 x 600 DPI)



Three plots demonstrating the bivariate relationship between the outcomes of interest and mean pain. Plot A shows decreased length of stays with increased mean pain levels. Plot B and Plot C show that, on average, those who expired at 30 days and 1 year marks experienced lower in hospital pain levels than those who did not expire.

114x99mm (600 x 600 DPI)

Regression description: multivar_linear model using mean_pain and ventdur<=24

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Number of Observations Read	844
Number of Observations Used	844

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	11888	2377.65640	70.77	<.0001
Error	838	28155	33.59817		
Corrected Total	843	40044			

Root MSE	5.79639	R-Square	0.2969
Dependent Mean	8.58776	Adj R-Sq	0.2927
Coeff Var	67.49599		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	6.74721	191633.70464	3.96	<.0001
mean_pain	1	-0.91580	0.12415	-7.38	<.0001
male	1	-1.78286	0.51402	-3.47	0.0006
age	1	0.00471	0.02021	0.23	0.8160
e_score	1	1.61599	0.12331	13.10	<.0001
oasis	1	0.09119	0.03159	2.89	0.0040

Regression description: multivar_linear model using mean_pain and ventdur<=24

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Regression description: multivar_linear model using med_pain and ventdur<=24

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Number of Observations Read	844
Number of Observations Used	844

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	11808	2361.65507	70.09	<.0001
Error	838	28235	33.69364		
Corrected Total	843	40044			

Root MSE	5.80462	R-Square	0.2949
Dependent Mean	8.58776	Adj R-Sq	0.2907
Coeff Var	67.59182		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	5.51742	1.66209	3.32	0.0009
med_pain	1	-0.69605	0.09657	-7.21	<.0001
male	1	-1.75771	0.51489	-3.41	0.0007
age	1	0.01249	0.02011	0.62	0.5346
e_score	1	1.62356	0.12339	13.16	<.0001
oasis	1	0.08689	0.03159	2.75	0.0061

Regression description: multivar_linear model using med_pain and ventdur<=24

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Regression description: multivar_linear model using max_pain and ventdur<=24

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Number of Observations Read	844
Number of Observations Used	844

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	10158	2031.60013	56.97	<.0001
Error	838	29886	35.66294		
Corrected Total	843	40044			

Root MSE	5.97185	R-Square	0.2537
Dependent Mean	8.58776	Adj R-Sq	0.2492
Coeff Var	69.53905		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.28286	1.84430	0.70	0.4869
max_pain	1	0.14781	0.08819	1.68	0.0941
male	1	-1.91709	0.52929	-3.62	0.0003
age	1	0.03550	0.02098	1.69	0.0910
e_score	1	1.79329	0.12447	14.41	<.0001
oasis	1	0.06871	0.03250	2.11	0.0348

Regression description: multivar_linear model using max_pain and ventdur<=24

The REG Procedure

Model: MODEL1
Dependent Variable: hosp_los

Regression description: multivar_linear model using cat_pain and ventdur<=24

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Number of Observations Read	844
Number of Observations Used	844

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	11727	2345.48240	69.41	<.0001
Error	838	28316	33.79014		
Corrected Total	843	40044			

Root MSE	5.81293	R-Square	0.2929
Dependent Mean	8.58776	Adj R-Sq	0.2886
Coeff Var	67.68854		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	6.47649	1.70556	3.80	0.0002
cat_pain	1	-2.26964	0.32289	-7.03	<.0001
male	1	-1.78270	0.51551	-3.46	0.0006
age	1	0.00679	0.02025	0.34	0.7376
e_score	1	1.62244	0.12372	13.11	<.0001
oasis	1	0.09063	0.03168	2.86	0.0043

Regression description: multivar_linear model using cat_pain and ventdur<=24

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Regression description: multivar_logistic_30day model using mean_pain and ventdur<=24

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEST2
Response Variable	X30_day
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	844
Number of Observations Used	844

Response Profile		
Ordered Value	X30_day	Total Frequency
1	1	21
2	0	823

Probability modeled is X30_day='1'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	198.606	149.641
SC	203.344	178.070
-2 Log L	196.606	137.641

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	58.9644	5	<.0001
Score	72.0933	5	<.0001
Wald	40.8033	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-7.3196	2.3007	10.1217	0.0015
mean_pain	1	-0.7830	0.2078	14.1967	0.0002
age	1	0.0268	0.0255	1.0995	0.2944
male	1	0.5256	0.5659	0.8627	0.3530
e_score	1	0.4041	0.1115	13.1357	0.0003
oasis	1	0.0553	0.0356	2.4204	0.1198

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
mean_pain	0.457	0.304	0.687
age	1.027	0.977	1.080
male	1.692	0.558	5.128
e_score	1.498	1.204	1.864
oasis	1.057	0.986	1.133

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	91.1	Somers' D	0.821
Percent Discordant	8.9	Gamma	0.821

Percent Tied	0.0	Tau-a	0.040
Pairs	17283	c	0.911

Partition for the Hosmer and Lemeshow Test					
		X30_day = 1		X30_day = 0	
Group	Total	Observed	Expected	Observed	Expected
1	84	0	0.04	84	83.96
2	84	0	0.09	84	83.91
3	84	0	0.15	84	83.85
4	84	0	0.24	84	83.76
5	84	1	0.36	83	83.64
6	84	0	0.54	84	83.46
7	84	0	0.84	84	83.16
8	84	3	1.45	81	82.55
9	84	3	3.13	81	80.87
10	88	14	14.16	74	73.84

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
4.7470	8	0.7842

Regression description: multivar_logistic_30day model using med_pain and ventdur<=24

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEST2
Response Variable	X30_day
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	844

Number of Observations Used	844

Response Profile		
Ordered Value	X30_day	Total Frequency
1	1	21
2	0	823

Probability modeled is X30_day='1'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	198.606	158.742
SC	203.344	187.171
-2 Log L	196.606	146.742

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	49.8631	5	<.0001
Score	63.0052	5	<.0001
Wald	38.4675	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-8.9316	2.1971	16.5261	<.0001
med_pain	1	-0.4474	0.1612	7.7053	0.0055
age	1	0.0377	0.0253	2.2238	0.1359
male	1	0.5085	0.5589	0.8276	0.3630

e_score	1	0.4428	0.1083	16.7092	<.0001
oasis	1	0.0548	0.0348	2.4831	0.1151

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
med_pain	0.639	0.466	0.877
age	1.038	0.988	1.091
male	1.663	0.556	4.973
e_score	1.557	1.259	1.925
oasis	1.056	0.987	1.131

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	88.4	Somers' D	0.768
Percent Discordant	11.6	Gamma	0.768
Percent Tied	0.0	Tau-a	0.037
Pairs	17283	c	0.884

Partition for the Hosmer and Lemeshow Test					
		X30_day = 1		X30_day = 0	
Group	Total	Observed	Expected	Observed	Expected
1	84	0	0.07	84	83.93
2	84	0	0.15	84	83.85
3	84	0	0.23	84	83.77
4	84	0	0.33	84	83.67
5	84	1	0.49	83	83.51
6	84	0	0.69	84	83.31
7	84	3	1.05	81	82.95
8	84	1	1.70	83	82.30
9	84	3	3.37	81	80.63
10	88	13	12.91	75	75.09

Hosmer and Lemeshow Goodness-of-Fit

Test		
Chi-Square	DF	Pr > ChiSq
6.0151	8	0.6455

Regression description: multivar_logistic_30day model using max_pain and ventdur<=24

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEST2
Response Variable	X30_day
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	844
Number of Observations Used	844

Response Profile		
Ordered Value	X30_day	Total Frequency
1	1	21
2	0	823

Probability modeled is X30_day='1'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	198.606	162.636

SC	203.344	191.065
-2 Log L	196.606	150.636

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	45.9693	5	<.0001
Score	61.5400	5	<.0001
Wald	39.1926	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-8.0597	2.2457	12.8804	0.0003
max_pain	1	-0.2081	0.0806	6.6669	0.0098
age	1	0.0276	0.0267	1.0721	0.3005
male	1	0.2187	0.5421	0.1628	0.6866
e_score	1	0.5779	0.1088	28.2346	<.0001
oasis	1	0.0587	0.0344	2.9173	0.0876

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
max_pain	0.812	0.693	0.951
age	1.028	0.976	1.083
male	1.244	0.430	3.601
e_score	1.782	1.440	2.206
oasis	1.060	0.991	1.134

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	86.8	Somers' D	0.736
Percent Discordant	13.2	Gamma	0.736
Percent Tied	0.0	Tau-a	0.036
Pairs	17283	c	0.868

Partition for the Hosmer and Lemeshow Test					
		X30_day = 1		X30_day = 0	
Group	Total	Observed	Expected	Observed	Expected
1	84	0	0.14	84	83.86
2	84	0	0.24	84	83.76
3	84	1	0.36	83	83.64
4	84	1	0.47	83	83.53
5	84	0	0.62	84	83.38
6	84	0	0.84	84	83.16
7	84	0	1.17	84	82.83
8	84	2	1.75	82	82.25
9	84	4	3.21	80	80.79
10	88	13	12.20	75	75.80

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
5.0983	8	0.7470

Regression description: multivar_logistic_30day model using cat_pain and ventdur<=24

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEST2
Response Variable	X30_day
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	844
Number of Observations Used	844

Response Profile		
Ordered Value	X30_day	Total Frequency
1	1	21
2	0	823

Probability modeled is X30_day='1'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	198.606	154.226
SC	203.344	182.654
-2 Log L	196.606	142.226

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	54.3800	5	<.0001
Score	69.7089	5	<.0001
Wald	42.5015	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-7.9103	2.2412	12.4572	0.0004
cat_pain	1	-1.5417	0.4355	12.5328	0.0004
age	1	0.0298	0.0255	1.3654	0.2426
male	1	0.5105	0.5639	0.8196	0.3653
e_score	1	0.4277	0.1127	14.4131	0.0001
oasis	1	0.0560	0.0345	2.6370	0.1044

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
cat_pain	0.214	0.091	0.502
age	1.030	0.980	1.083
male	1.666	0.552	5.031
e_score	1.534	1.230	1.913
oasis	1.058	0.988	1.132

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	90.7	Somers' D	0.814
Percent Discordant	9.3	Gamma	0.814
Percent Tied	0.0	Tau-a	0.040
Pairs	17283	c	0.907

Partition for the Hosmer and Lemeshow Test					
		X30_day = 1		X30_day = 0	
Group	Total	Observed	Expected	Observed	Expected
1	84	0	0.07	84	83.93
2	85	0	0.14	85	84.86
3	84	0	0.23	84	83.77
4	84	0	0.37	84	83.63
5	84	0	0.52	84	83.48
6	84	1	0.69	83	83.31
7	84	0	1.02	84	82.98
8	84	2	1.51	82	82.49
9	84	3	2.93	81	81.07
10	87	15	13.52	72	73.48

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
2.8512	8	0.9433

Regression description: multivar_logistic_1yr model using mean_pain and ventdur<=24

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEST2
Response Variable	X1_yr
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	844
Number of Observations Used	844

Response Profile		
Ordered Value	X1_yr	Total Frequency
1	1	46
2	0	798

Probability modeled is X1_yr='1'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	359.121	282.785
SC	363.859	311.214
-2 Log L	357.121	270.785

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	86.3361	5	<.0001
Score	100.3926	5	<.0001
Wald	64.9324	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-9.2928	1.6644	31.1714	<.0001
mean_pain	1	-0.3430	0.1105	9.6411	0.0019
age	1	0.0599	0.0191	9.8411	0.0017
male	1	0.3160	0.3883	0.6622	0.4158
e_score	1	0.4610	0.0834	30.5477	<.0001
oasis	1	0.0496	0.0243	4.1861	0.0408

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
mean_pain	0.710	0.571	0.881
age	1.062	1.023	1.102
male	1.372	0.641	2.936
e_score	1.586	1.347	1.867
oasis	1.051	1.002	1.102

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	85.2	**Somers' D**	0.704
Percent Discordant	14.8	**Gamma**	0.704
Percent Tied	0.0	**Tau-a**	0.073
Pairs	36708	**c**	0.852

Partition for the Hosmer and Lemeshow Test			
Group	**Total**	**X1_yr = 1**	**X1_yr = 0**

		Observed	Expected	Observed	Expected
1	84	0	0.23	84	83.77
2	84	0	0.49	84	83.51
3	84	0	0.78	84	83.22
4	84	3	1.11	81	82.89
5	84	2	1.50	82	82.50
6	84	1	2.04	83	81.96
7	84	3	2.94	81	81.06
8	84	3	4.67	81	79.33
9	84	10	7.86	74	76.14
10	88	24	24.36	64	63.64

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
6.7640	8	0.5623

Regression description: multivar_logistic_1yr model using med_pain and ventdur<=24

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEST2
Response Variable	X1_yr
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	844
Number of Observations Used	844

Response Profile		
Ordered Value	X1_yr	Total Frequency
1	1	46

2	0	798

Probability modeled is X1_yr='1'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	359.121	289.827
SC	363.859	318.256
-2 Log L	357.121	277.827

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	79.2945	5	<.0001
Score	92.8046	5	<.0001
Wald	62.5204	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-10.0363	1.6198	38.3892	<.0001
med_pain	1	-0.1552	0.0835	3.4581	0.0629
age	1	0.0647	0.0190	11.5703	0.0007
male	1	0.2607	0.3844	0.4600	0.4976
e_score	1	0.4868	0.0828	34.5309	<.0001
oasis	1	0.0461	0.0239	3.7018	0.0544

Odds Ratio Estimates		
Effect	Point Estimate	95% Wald Confidence Limits

med_pain	0.856	0.727	1.008
age	1.067	1.028	1.107
male	1.298	0.611	2.757
e_score	1.627	1.383	1.914
oasis	1.047	0.999	1.097

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	83.6	Somers' D	0.672
Percent Discordant	16.4	Gamma	0.672
Percent Tied	0.0	Tau-a	0.069
Pairs	36708	c	0.836

Partition for the Hosmer and Lemeshow Test					
		X1_yr = 1		X1_yr = 0	
Group	Total	Observed	Expected	Observed	Expected
1	84	0	0.28	84	83.72
2	84	1	0.60	83	83.40
3	84	0	0.89	84	83.11
4	84	2	1.23	82	82.77
5	84	3	1.63	81	82.37
6	84	1	2.24	83	81.76
7	84	3	3.24	81	80.76
8	84	2	4.90	82	79.10
9	84	10	7.45	74	76.55
10	88	24	23.55	64	64.45

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
6.6165	8	0.5785

Regression description: multivar_logistic_1yr model using max_pain and ventdur<=24

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEST2
Response Variable	X1_yr
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	844
Number of Observations Used	844

Response Profile		
Ordered Value	X1_yr	Total Frequency
1	1	46
2	0	798

Probability modeled is X1_yr='1'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	359.121	289.427
SC	363.859	317.856
-2 Log L	357.121	277.427

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	79.6942	5	<.0001

Score	93.2270	5	<.0001
Wald	62.4345	5	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-9.4442	1.6770	31.7153	<.0001
max_pain	1	-0.1205	0.0591	4.1619	0.0413
age	1	0.0596	0.0194	9.3917	0.0022
male	1	0.1720	0.3776	0.2074	0.6488
e_score	1	0.5437	0.0822	43.7465	<.0001
oasis	1	0.0487	0.0241	4.0913	0.0431

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
max_pain	0.887	0.790	0.995
age	1.061	1.022	1.103
male	1.188	0.567	2.490
e_score	1.722	1.466	2.023
oasis	1.050	1.002	1.101

Association of Predicted Probabilities and Observed Responses

Percent Concordant	83.1	**Somers' D**	0.663
Percent Discordant	16.9	**Gamma**	0.663
Percent Tied	0.0	**Tau-a**	0.068
Pairs	36708	**c**	0.831

Partition for the Hosmer and Lemeshow Test

Group	Total	X1_yr = 1		X1_yr = 0	
		Observed	Expected	Observed	Expected
1	84	1	0.29	83	83.71
2	84	0	0.58	84	83.42

3	85	2	0.90	83	84.10
4	84	1	1.23	83	82.77
5	84	0	1.68	84	82.32
6	84	4	2.26	80	81.74
7	84	2	3.16	82	80.84
8	84	4	4.75	80	79.25
9	84	8	7.94	76	76.06
10	87	24	23.21	63	63.79

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
7.4004	8	0.4941

Regression description: multivar_logistic_1yr model using cat_pain and ventdur<=24

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEST2
Response Variable	X1_yr
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	844
Number of Observations Used	844

Response Profile		
Ordered Value	X1_yr	Total Frequency
1	1	46
2	0	798

Probability modeled is X1_yr='1'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	359.121	284.013
SC	363.859	312.442
-2 Log L	357.121	272.013

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	85.1076	5	<.0001
Score	99.4422	5	<.0001
Wald	64.6025	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-9.4525	1.6507	32.7923	<.0001
cat_pain	1	-0.7994	0.2680	8.8971	0.0029
age	1	0.0605	0.0190	10.1350	0.0015
male	1	0.3156	0.3878	0.6624	0.4157
e_score	1	0.4689	0.0836	31.4847	<.0001
oasis	1	0.0501	0.0241	4.3314	0.0374

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
cat_pain	0.450	0.266	0.760
age	1.062	1.024	1.103
male	1.371	0.641	2.932

e_score	1.598	1.357	1.883
oasis	1.051	1.003	1.102

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	85.1	Somers' D	0.702
Percent Discordant	14.9	Gamma	0.702
Percent Tied	0.0	Tau-a	0.072
Pairs	36708	c	0.851

Partition for the Hosmer and Lemeshow Test					
		X1_yr = 1		X1_yr = 0	
Group	Total	Observed	Expected	Observed	Expected
1	84	0	0.25	84	83.75
2	84	0	0.52	84	83.48
3	84	0	0.78	84	83.22
4	84	4	1.12	80	82.88
5	84	1	1.55	83	82.45
6	84	1	2.11	83	81.89
7	84	1	3.14	83	80.86
8	84	5	4.69	79	79.31
9	84	10	7.81	74	76.19
10	88	24	24.03	64	63.97

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
12.1299	8	0.1455

Regression description: multivar_linear model using mean_pain

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Number of Observations Read	1889

Number of Observations Used	1889

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	13989	2797.85361	96.07	<.0001
Error	1883	54838	29.12252		
Corrected Total	1888	68827			

Root MSE	5.39653	R-Square	0.2033
Dependent Mean	9.05966	Adj R-Sq	0.2011
Coeff Var	59.56654		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	6.13876	1.07024	5.74	<.0001
mean_pain	1	-0.70945	0.08044	-8.82	<.0001
male	1	-1.08190	0.29793	-3.63	0.0003
age	1	0.01732	0.01270	1.36	0.1728
e_score	1	1.13959	0.07364	15.47	<.0001
oasis	1	0.07134	0.01895	3.76	0.0002

Regression description: multivar_linear model using mean_pain

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Regression description: multivar_logistic_30day model using mean_pain

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEST2
Response Variable	X30_day
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1889
Number of Observations Used	1889

Response Profile		
Ordered Value	X30_day	Total Frequency
1	1	38
2	0	1851

Probability modeled is X30_day='1'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	374.103	324.079
SC	379.647	357.342
-2 Log L	372.103	312.079

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	60.0235	5	<.0001
Score	66.7408	5	<.0001
Wald	53.0781	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-6.5574	1.5435	18.0499	<.0001
mean_pain	1	-0.5241	0.1328	15.5838	<.0001
age	1	0.0188	0.0180	1.0847	0.2977
male	1	-0.0844	0.3616	0.0545	0.8154
e_score	1	0.3246	0.0785	17.0945	<.0001
oasis	1	0.0482	0.0244	3.9116	0.0480

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
mean_pain	0.592	0.456	0.768
age	1.019	0.984	1.056
male	0.919	0.452	1.867
e_score	1.384	1.186	1.614
oasis	1.049	1.000	1.101

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	81.1	Somers' D	0.622
Percent Discordant	18.9	Gamma	0.622
Percent Tied	0.0	Tau-a	0.025
Pairs	70338	c	0.811

Partition for the Hosmer and Lemeshow Test					
		X30_day = 1		X30_day = 0	
Group	Total	Observed	Expected	Observed	Expected
1	190	1	0.26	189	189.74
2	189	0	0.51	189	188.49
3	189	0	0.76	189	188.24
4	189	2	1.01	187	187.99

5	189	1	1.39	188	187.61
6	189	0	1.90	189	187.10
7	189	8	2.71	181	186.29
8	189	2	4.04	187	184.96
9	189	4	6.87	185	182.13
10	187	20	18.56	167	168.44

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
19.2358	8	0.0136

Regression description: multivar_logistic_1yr model using mean_pain

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEST2
Response Variable	X1_yr
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1889
Number of Observations Used	1889

Response Profile		
Ordered Value	X1_yr	Total Frequency
1	1	104
2	0	1785

Probability modeled is X1_yr='1'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	807.244	702.254
SC	812.788	735.517
-2 Log L	805.244	690.254

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	114.9898	5	<.0001
Score	129.4134	5	<.0001
Wald	104.6689	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-7.8571	0.9995	61.7908	<.0001
mean_pain	1	-0.1076	0.0684	2.4728	0.1158
age	1	0.0413	0.0118	12.2649	0.0005
male	1	-0.0289	0.2302	0.0158	0.9000
e_score	1	0.4230	0.0523	65.3771	<.0001
oasis	1	0.0366	0.0150	5.9631	0.0146

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
mean_pain	0.898	0.785	1.027
age	1.042	1.018	1.066
male	0.971	0.619	1.525
e_score	1.527	1.378	1.691
oasis	1.037	1.007	1.068

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	78.6	Somers' D	0.572
Percent Discordant	21.4	Gamma	0.572
Percent Tied	0.0	Tau-a	0.060
Pairs	185640	c	0.786

Partition for the Hosmer and Lemeshow Test					
		X1_yr = 1		X1_yr = 0	
Group	Total	Observed	Expected	Observed	Expected
1	189	0	1.37	189	187.63
2	189	3	2.35	186	186.65
3	189	2	3.27	187	185.73
4	189	6	4.21	183	184.79
5	189	4	5.28	185	183.72
6	189	5	6.86	184	182.14
7	189	11	8.87	178	180.13
8	189	13	11.86	176	177.14
9	189	14	18.27	175	170.73
10	188	46	41.67	142	146.33

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
6.0183	8	0.6452

Regression description: multivar_linear model using cat_pain

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Number of Observations Read	1889
Number of Observations Used	1889

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	13724	2744.72524	93.79	<.0001
Error	1883	55103	29.26359		
Corrected Total	1888	68827			

Root MSE	5.40958	R-Square	0.1994
Dependent Mean	9.05966	Adj R-Sq	0.1973
Coeff Var	59.71064		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	5.70144	1.06297	5.36	<.0001
cat_pain	1	-1.70596	0.20636	-8.27	<.0001
male	1	-1.04945	0.29877	-3.51	0.0005
age	1	0.02149	0.01266	1.70	0.0900
e_score	1	1.14537	0.07384	15.51	<.0001
oasis	1	0.07046	0.01900	3.71	0.0002

Regression description: multivar_linear model using cat_pain

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Regression description: multivar_logistic_30day model using cat_pain

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEST2

Response Variable	X30_day
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1889
Number of Observations Used	1889

Response Profile

Ordered Value	X30_day	Total Frequency
1	1	38
2	0	1851

Probability modeled is X30_day='1'.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	374.103	327.406
SC	379.647	360.669
-2 Log L	372.103	315.406

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	56.6967	5	<.0001
Score	64.5808	5	<.0001
Wald	53.1142	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-6.9727	1.5104	21.3116	<.0001
cat_pain	1	-1.1138	0.2957	14.1909	0.0002
age	1	0.0213	0.0180	1.3982	0.2370
male	1	-0.0529	0.3617	0.0214	0.8836
e_score	1	0.3400	0.0788	18.6130	<.0001
oasis	1	0.0482	0.0243	3.9417	0.0471

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
cat_pain	0.328	0.184	0.586
age	1.022	0.986	1.058
male	0.948	0.467	1.927
e_score	1.405	1.204	1.640
oasis	1.049	1.001	1.100

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	80.4	Somers' D	0.607
Percent Discordant	19.6	Gamma	0.607
Percent Tied	0.0	Tau-a	0.024
Pairs	70338	c	0.804

Partition for the Hosmer and Lemeshow Test					
		X30_day = 1		X30_day = 0	
Group	Total	Observed	Expected	Observed	Expected
1	189	0	0.31	189	188.69
2	189	1	0.60	188	188.40
3	189	1	0.90	188	188.10
4	189	1	1.21	188	187.79
5	189	3	1.57	186	187.43
6	189	0	2.04	189	186.96

7	189	5	2.67	184	186.33
8	189	3	3.91	186	185.09
9	189	7	6.63	182	182.37
10	188	17	18.17	171	169.83

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
6.3800	8	0.6047

Regression description: multivar_logistic_1yr model using cat_pain

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEST2
Response Variable	X1_yr
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1889
Number of Observations Used	1889

Response Profile		
Ordered Value	X1_yr	Total Frequency
1	1	104
2	0	1785

Probability modeled is X1_yr='1'.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	807.244	701.692
SC	812.788	734.955
-2 Log L	805.244	689.692

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	115.5518	5	<.0001
Score	129.8715	5	<.0001
Wald	104.8178	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-7.8363	0.9896	62.7055	<.0001
cat_pain	1	-0.3017	0.1725	3.0581	0.0803
age	1	0.0411	0.0117	12.3074	0.0005
male	1	-0.0158	0.2305	0.0047	0.9454
e_score	1	0.4232	0.0522	65.6557	<.0001
oasis	1	0.0367	0.0150	5.9893	0.0144

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
cat_pain	0.740	0.527	1.037
age	1.042	1.018	1.066
male	0.984	0.626	1.547
e_score	1.527	1.378	1.691
oasis	1.037	1.007	1.068

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	78.7	Somers' D	0.574
Percent Discordant	21.3	Gamma	0.574
Percent Tied	0.0	Tau-a	0.060
Pairs	185640	c	0.787

Partition for the Hosmer and Lemeshow Test					
		X1_yr = 1		X1_yr = 0	
Group	Total	Observed	Expected	Observed	Expected
1	189	1	1.36	188	187.64
2	189	1	2.33	188	186.67
3	189	5	3.23	184	185.77
4	189	3	4.18	186	184.82
5	189	6	5.32	183	183.68
6	189	6	6.81	183	182.19
7	189	10	8.84	179	180.16
8	189	12	11.88	177	177.12
9	189	16	18.34	173	170.66
10	188	44	41.71	144	146.29

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
3.0321	8	0.9323

Regression description: multivar_linear model using mean_pain and hospital_expire_flag = 0

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Number of Observations Read	1867
Number of Observations Used	1867

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	12848	2569.60206	91.82	<.0001
Error	1861	52078	27.98408		
Corrected Total	1866	64926			

Root MSE	5.29000	R-Square	0.1979
Dependent Mean	9.01968	Adj R-Sq	0.1957
Coeff Var	58.64951		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	6.06474	1.05804	5.73	<.0001
mean_pain	1	-0.70050	0.07969	-8.79	<.0001
male	1	-0.95798	0.29392	-3.26	0.0011
age	1	0.01991	0.01256	1.58	0.1132
e_score	1	1.11301	0.07288	15.27	<.0001
oasis	1	0.06609	0.01871	3.53	0.0004

Regression description: multivar_linear model using mean_pain and hospital_expire_flag = 0

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Regression description: multivar_logistic_30day model using mean_pain and hospital_expire_flag = 0

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEST2
Response Variable	X30_day

Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1867
Number of Observations Used	1867

Response Profile		
Ordered Value	X30_day	Total Frequency
1	1	16
2	0	1851

Probability modeled is X30_day='1'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	186.166	172.675
SC	191.699	205.868
-2 Log L	184.166	160.675

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	23.4914	5	0.0003
Score	27.1882	5	<.0001
Wald	23.1706	5	0.0003

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-10.2879	2.4688	17.3652	<.0001
mean_pain	1	-0.2196	0.1777	1.5273	0.2165
age	1	0.0500	0.0295	2.8652	0.0905
male	1	0.0170	0.5518	0.0009	0.9755
e_score	1	0.3972	0.1162	11.6884	0.0006
oasis	1	0.0394	0.0358	1.2124	0.2709

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
mean_pain	0.803	0.567	1.137
age	1.051	0.992	1.114
male	1.017	0.345	2.999
e_score	1.488	1.185	1.868
oasis	1.040	0.970	1.116

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	79.7	Somers' D	0.593
Percent Discordant	20.3	Gamma	0.593
Percent Tied	0.0	Tau-a	0.010
Pairs	29616	c	0.797

Partition for the Hosmer and Lemeshow Test					
		X30_day = 1		X30_day = 0	
Group	Total	Observed	Expected	Observed	Expected
1	187	0	0.13	187	186.87
2	187	0	0.26	187	186.74
3	187	0	0.38	187	186.62
4	187	2	0.50	185	186.50
5	188	1	0.65	187	187.35
6	187	0	0.88	187	186.12
7	187	2	1.21	185	185.79

8	187	2	1.71	185	185.29
9	187	0	2.68	187	184.32
10	183	9	7.60	174	175.40

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
9.8622	8	0.2748

Regression description: multivar_logistic_1yr model using mean_pain and hospital_expire_flag = 0

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEST2
Response Variable	X1_yr
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1867
Number of Observations Used	1867

Response Profile		
Ordered Value	X1_yr	Total Frequency
1	1	82
2	0	1785

Probability modeled is X1_yr='1'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	674.905	591.678
SC	680.437	624.870
-2 Log L	672.905	579.678

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	93.2272	5	<.0001
Score	104.3845	5	<.0001
Wald	86.3856	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-9.1906	1.1351	65.5626	<.0001
mean_pain	1	0.0267	0.0738	0.1307	0.7177
age	1	0.0530	0.0135	15.5081	<.0001
male	1	-0.00300	0.2564	0.0001	0.9907
e_score	1	0.4467	0.0577	59.9324	<.0001
oasis	1	0.0309	0.0165	3.5163	0.0608

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
mean_pain	1.027	0.889	1.187
age	1.054	1.027	1.083
male	0.997	0.603	1.648
e_score	1.563	1.396	1.750
oasis	1.031	0.999	1.065

Association of Predicted Probabilities and

Observed Responses			
Percent Concordant	78.8	Somers' D	0.575
Percent Discordant	21.2	Gamma	0.575
Percent Tied	0.0	Tau-a	0.048
Pairs	146370	c	0.788

Partition for the Hosmer and Lemeshow Test					
		X1_yr = 1		X1_yr = 0	
Group	Total	Observed	Expected	Observed	Expected
1	187	1	0.99	186	186.01
2	187	1	1.75	186	185.25
3	187	2	2.46	185	184.54
4	187	3	3.26	184	183.74
5	187	6	4.20	181	182.80
6	187	1	5.35	186	181.65
7	187	11	7.02	176	179.98
8	187	13	9.26	174	177.74
9	187	8	14.23	179	172.77
10	184	36	33.49	148	150.51

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
11.9787	8	0.1522

Regression description: multivar_linear model using cat_pain and hospital_expire_flag = 0

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Number of Observations Read	1867
Number of Observations Used	1867

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	12579	2515.76531	89.44	<.0001
Error	1861	52348	28.12873		
Corrected Total	1866	64926			

Root MSE	5.30365	R-Square	0.1937
Dependent Mean	9.01968	Adj R-Sq	0.1916
Coeff Var	58.80089		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	5.63343	1.05160	5.36	<.0001
cat_pain	1	-1.68014	0.20479	-8.20	<.0001
male	1	-0.92734	0.29477	-3.15	0.0017
age	1	0.02397	0.01253	1.91	0.0559
e_score	1	1.11908	0.07308	15.31	<.0001
oasis	1	0.06514	0.01875	3.47	0.0005

Regression description: multivar_linear model using cat_pain and hospital_expire_flag = 0

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Regression description: multivar_logistic_30day model using cat_pain and hospital_expire_flag = 0

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEST2
Response Variable	X30_day

Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1867
Number of Observations Used	1867

Response Profile

Ordered Value	X30_day	Total Frequency
1	1	16
2	0	1851

Probability modeled is X30_day='1'.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	186.166	173.641
SC	191.699	206.834
-2 Log L	184.166	161.641

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	22.5254	5	0.0004
Score	26.1846	5	<.0001
Wald	22.6468	5	0.0004

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-10.7032	2.4542	19.0199	<.0001
cat_pain	1	-0.3439	0.4230	0.6607	0.4163
age	1	0.0531	0.0296	3.2079	0.0733
male	1	0.0324	0.5518	0.0035	0.9531
e_score	1	0.4126	0.1162	12.6022	0.0004
oasis	1	0.0375	0.0354	1.1222	0.2894

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
cat_pain	0.709	0.309	1.625
age	1.055	0.995	1.118
male	1.033	0.350	3.046
e_score	1.511	1.203	1.897
oasis	1.038	0.969	1.113

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	79.6	Somers' D	0.592
Percent Discordant	20.4	Gamma	0.592
Percent Tied	0.0	Tau-a	0.010
Pairs	29616	c	0.796

Partition for the Hosmer and Lemeshow Test					
		X30_day = 1		X30_day = 0	
Group	Total	Observed	Expected	Observed	Expected
1	187	0	0.15	187	186.85
2	188	0	0.28	188	187.72
3	188	1	0.40	187	187.60
4	187	1	0.53	186	186.47
5	187	0	0.69	187	186.31
6	187	1	0.92	186	186.08
7	187	2	1.23	185	185.77

8	187	1	1.72	186	185.28
9	187	2	2.66	185	184.34
10	182	8	7.42	174	174.58

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
3.4540	8	0.9027

Regression description: multivar_logistic_1yr model using cat_pain and hospital_expire_flag = 0

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEST2
Response Variable	X1_yr
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1867
Number of Observations Used	1867

Response Profile		
Ordered Value	X1_yr	Total Frequency
1	1	82
2	0	1785

Probability modeled is X1_yr='1'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	674.905	591.769
SC	680.437	624.962
-2 Log L	672.905	579.769

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	93.1360	5	<.0001
Score	104.3571	5	<.0001
Wald	86.3552	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-9.1205	1.1300	65.1401	<.0001
cat_pain	1	0.0376	0.1909	0.0389	0.8436
age	1	0.0524	0.0134	15.2784	<.0001
male	1	-0.00484	0.2565	0.0004	0.9850
e_score	1	0.4444	0.0574	59.9633	<.0001
oasis	1	0.0312	0.0165	3.5868	0.0582

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
cat_pain	1.038	0.714	1.509
age	1.054	1.026	1.082
male	0.995	0.602	1.645
e_score	1.560	1.394	1.745
oasis	1.032	0.999	1.066

Association of Predicted Probabilities and Observed Responses

Percent Concordant	78.8	Somers' D	0.576
Percent Discordant	21.2	Gamma	0.576
Percent Tied	0.0	Tau-a	0.048
Pairs	146370	c	0.788

Partition for the Hosmer and Lemeshow Test

		X1_yr = 1		X1_yr = 0	
Group	Total	Observed	Expected	Observed	Expected
1	187	1	1.00	186	186.00
2	187	1	1.75	186	185.25
3	187	2	2.47	185	184.53
4	187	3	3.26	184	183.74
5	187	6	4.20	181	182.80
6	187	1	5.36	186	181.64
7	187	9	7.01	178	179.99
8	187	15	9.27	172	177.73
9	187	9	14.22	178	172.78
10	184	35	33.48	149	150.52

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
11.3564	8	0.1823

Regression description: multivar_linear model using delirium and ventdur<=24

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Number of Observations Read	844
Number of Observations Used	844

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F

Model	5	12159	2431.79827	73.08	<.0001
Error	838	27885	33.27513		
Corrected Total	843	40044			

Root MSE	5.76846	**R-Square**	0.3036
Dependent Mean	8.58776	**Adj R-Sq**	0.2995
Coeff Var	67.17072		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	3.63479	1.60805	2.26	0.0241
delirium	1	5.56526	0.70035	7.95	<.0001
male	1	-1.94421	0.51127	-3.80	0.0002
age	1	0.01416	0.01994	0.71	0.4777
e_score	1	1.57040	0.12361	12.70	<.0001
oasis	1	0.06822	0.03134	2.18	0.0297

Regression description: multivar_linear model using delirium and ventdur<=24

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Regression description: multivar_logistic_30day model using delirium and ventdur<=24

The LOGISTIC Procedure

Model Information

Data Set	WORK.TEST2
Response Variable	X30_day
Number of Response Levels	2
Model	binary logit

Optimization Technique	Fisher's scoring

Number of Observations Read	844
Number of Observations Used	844

Response Profile		
Ordered Value	X30_day	Total Frequency
1	1	21
2	0	823

Probability modeled is X30_day='1'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	198.606	169.078
SC	203.344	197.507
-2 Log L	196.606	157.078

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	39.5271	5	<.0001
Score	54.3318	5	<.0001
Wald	36.6080	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq

Intercept	1	-9.8951	2.1345	21.4911	<.0001
delirium	1	0.0949	0.6062	0.0245	0.8756
age	1	0.0401	0.0258	2.4110	0.1205
male	1	0.2956	0.5460	0.2931	0.5882
e_score	1	0.5498	0.1148	22.9383	<.0001
oasis	1	0.0492	0.0330	2.2217	0.1361

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
delirium	1.100	0.335	3.607
age	1.041	0.990	1.095
male	1.344	0.461	3.919
e_score	1.733	1.384	2.170
oasis	1.050	0.985	1.121

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	83.4	**Somers' D**	0.668
Percent Discordant	16.6	**Gamma**	0.668
Percent Tied	0.0	**Tau-a**	0.032
Pairs	17283	**c**	0.834

Partition for the Hosmer and Lemeshow Test					
		X30_day = 1		**X30_day = 0**	
Group	**Total**	**Observed**	**Expected**	**Observed**	**Expected**
1	84	0	0.19	84	83.81
2	84	0	0.32	84	83.68
3	84	2	0.44	82	83.56
4	84	0	0.58	84	83.42
5	84	1	0.74	83	83.26
6	84	0	0.99	84	83.01
7	84	0	1.31	84	82.69
8	84	3	1.91	81	82.09
9	84	2	3.08	82	80.92

10	88	13	11.43	75	76.57

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	**DF**	**Pr > ChiSq**
10.4118	8	0.2373

Regression description: multivar_linear model using nausea and ventdur<=24

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Number of Observations Read	844
Number of Observations Used	844

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	10064	2012.88265	56.27	<.0001
Error	838	29979	35.77462		
Corrected Total	843	40044			

Root MSE	5.98119	**R-Square**	0.2513
Dependent Mean	8.58776	**Adj R-Sq**	0.2469
Coeff Var	69.64784		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	2.56872	1.66855	1.54	0.1241
nausea	1	0.28090	0.65482	0.43	0.6680
male	1	-1.88947	0.53247	-3.55	0.0004
age	1	0.02862	0.02059	1.39	0.1650
e_score	1	1.80894	0.12539	14.43	<.0001
oasis	1	0.07194	0.03249	2.21	0.0271

Regression description: multivar_linear model using nausea and ventdur<=24

The REG Procedure
Model: MODEL1
Dependent Variable: hosp_los

Regression description: multivar_logistic_30day model using nausea and ventdur<=24

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TEST2
Response Variable	X30_day
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	844
Number of Observations Used	844

Response Profile		
Ordered Value	X30_day	Total Frequency
1	1	21
2	0	823

Probability modeled is X30_day='1'.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	198.606	167.594
SC	203.344	196.023
-2 Log L	196.606	155.594

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	41.0115	5	<.0001
Score	54.4566	5	<.0001
Wald	37.0514	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-10.1946	2.1297	22.9145	<.0001
nausea	1	0.9052	0.6819	1.7622	0.1844
age	1	0.0398	0.0255	2.4339	0.1187
male	1	0.3894	0.5555	0.4915	0.4833
e_score	1	0.5825	0.1080	29.1014	<.0001
oasis	1	0.0506	0.0331	2.3426	0.1259

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
nausea	2.473	0.650	9.410
age	1.041	0.990	1.094
male	1.476	0.497	4.385
e_score	1.791	1.449	2.213
oasis	1.052	0.986	1.122

Association of Predicted Probabilities and Observed Responses

Percent Concordant	83.5	Somers' D	0.670
Percent Discordant	16.5	Gamma	0.670
Percent Tied	0.0	Tau-a	0.033
Pairs	17283	c	0.835

Partition for the Hosmer and Lemeshow Test

		X30_day = 1		X30_day = 0	
Group	Total	Observed	Expected	Observed	Expected
1	84	0	0.17	84	83.83
2	84	0	0.29	84	83.71
3	84	2	0.40	82	83.60
4	84	0	0.54	84	83.46
5	84	1	0.70	83	83.30
6	84	0	0.95	84	83.05
7	84	1	1.29	83	82.71
8	84	1	1.89	83	82.11
9	84	4	3.05	80	80.95
10	88	12	11.71	76	76.29

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
9.2376	8	0.3226

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*

Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any pre-specified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	3-4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	4
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4-5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*4-5*
Bias	9	Describe any efforts to address potential sources of bias	9
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6
		(*b*) Describe any methods used to examine subgroups and interactions	6
		(*c*) Explain how missing data were addressed	6
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	NA

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
		(*e*) Describe any sensitivity analyses	6
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	NA
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6-7, Table 1
		(b) Indicate number of participants with missing data for each variable of interest	NA
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	8, Table 2
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Table 2
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	7
Discussion			
Key results	18	Summarise key results with reference to study objectives	7
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	9
Generalisability	21	Discuss the generalisability (external validity) of the study results	9
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	13

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.
Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Counterintuitive Results From Observational Data: A Case Study and Discussion

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-026447.R2
Article Type:	Research
Date Submitted by the Author:	21-Feb-2019
Complete List of Authors:	Doty, Erik; Harvard Medical School Department of Biomedical Informatics Stone, David; University of Virginia School of Medicine; Massachusetts Institute of Technology, Laboratory for Computational Physiology McCague, Ned; Massachusetts Institute of Technology, Laboratory for Computational Physiology; Kyruus, Inc Celi, Leo Anthony; Massachusetts Institute of Technology,
Primary Subject Heading:	Intensive care
Secondary Subject Heading:	Health informatics
Keywords:	Pain, Mortality, Length of Stay, Big Data, Counterintuitive results

Counterintuitive Results From Observational Data: A Case Study and Discussion

Erik Doty[1], DO, MBI; David J. Stone[2,5], MD; Ned McCague[3,5], MPH; Leo Anthony Celi[4,5], MD, MS, MPH

Corresponding author: Leo Anthony Celi

Email: lceli@bidmc.harvard.edu

Phone: (617) 253-7937

1. *Harvard Medical School Department of Biomedical Informatics, Boston, MA*
2. *University of Virginia School of Medicine, Charlottesville, VA*
3. *Kyruus, Inc., Boston, MA*
4. *Beth Israel Deaconess Medical Center, Boston, MA*
5. *MIT Laboratory for Computational Physiology, Cambridge, MA*

Counterintuitive Results from Observational Data: A Case Study and Discussion

E. Doty, D.J. Stone, N. McCague, L.A. Celi

Keywords: Pain, mortality, length of stay

Abstract

Objective: To explore the issue of counterintuitive data via analysis of a representative case in which the data obtained was unexpected and inconsistent with current knowledge. We then discuss the issue of counterintuitive data while developing a framework for approaching such findings.

Design: The case study is a retrospective analysis of a cohort of Coronary Artery Bypass Graft (CABG) patients. Regression was used to examine the association between perceived pain in the ICU and selected outcomes.

Setting: MIMIC-III, a publicly available, deidentified critical care patient database.

Participants: 844 adult patients from the database who underwent CABG surgery and were extubated within 24 hours after ICU admission.

Outcomes: 30-day mortality, 1-year mortality, and hospital length of stay (LOS).

Results: Increased pain levels were found to be significantly associated with reduced mortality at 30 days and 1-year, and shorter hospital LOS. A one-point increase in mean pain level was found to be associated with a reduction in the odds of 30-day and 1-year mortality by a factor of 0.457 (95%CI 0.304-0.687, p< 0.01) and 0.710 (95%CI 0.571 - 0.881, p< 0.01) respectively, and a 0.916 (95%CI (-1.159, -0.673), p < 0.01) day decrease in hospital LOS.

Conclusion: The finding of an association between increased pain and improved outcomes was unexpected and clinically counterintuitive. In an increasingly digitized age of medical big data, such results are likely to become more common. The reliability of such counterintuitive results must be carefully examined: We suggest several issues to consider in this analytic process. If the data is determined to be valid, consideration must then be made towards alternative explanations for the counterintuitive results observed. Such results may in fact indicate that current clinical knowledge is incomplete or not have been firmly based on empirical evidence, and function to inspire further research into the factors involved.

Strengths and limitations of this study

- Large sample size with minimal covariate data missing.
- Multiple regression models with multiple sensitivity analyses.

- High internal validity shown by use of falsification hypothesis testing.
- Lack of oral analgesic data.
- Recognizing that correlation does not equal causation and further work is needed to confirm case results.

Introduction

What do we mean by counterintuitive data? It is data that presents unexpected results that may clash with common sense or what has been previously published and accepted by the medical community. In practice, clinicians have long dealt with such results in individual bits but have had the vast advantage of being able to examine the concurrent state of the patient and react in real time by repeating a lab test or tracking ongoing monitor data. These responses function to identify the prior result as a non-repeatable error, or as a genuine anomaly. However, this approach is not applicable to the context of retrospective data analysis. Furthermore, the counterintuitive data revealed in such analyses is likely to be more involved than a single aberrant lab or vital sign value. In today's data driven healthcare system, retrospective data analyses are becoming more and more common. We can therefore logically expect to encounter a greater incidence and variety of counterintuitive values and results that are impossible to confirm by repetition, difficult to confirm or deny by context, but still require interpretation.

The question then becomes how best to approach such results? Are they incorrect simply because they weren't what was expected? And was the expectation itself based on subjective assumptions or objective conclusions? When our prior expectations are not met, are we dealing with truly faulty data, or do our expectations need to be reset by what are reliable, but counterintuitive, results. For example, we have learned that intensive care practices common in the past such as large tidal volume ventilation, the use of pulmonary artery catheters, and the use of lidocaine infusions in myocardial infarction led to no benefit or injury.[1-3] Were these unexpected negative outcomes initially missed because outcomes data was not being carefully analyzed, or perhaps ignored or interpreted as counterintuitive to the level of unbelievability? How can the situation be dissected retrospectively so that counterintuitive data can be identified as truly spurious versus simply not being consistent with our prior experience which may itself be faulty and require data driven correction?

In this paper, we explore a case in which the results contradicted previous reports and our clinical expectations. Using the Medical Information Mart for Intensive Care-III (MIMIC-III), a critical care database that was developed and maintained by the Laboratory for Computational Physiology at the Massachusetts Institute of Technology[4], we retrospectively selected a cohort of patients that underwent a coronary artery bypass graft (CABG) procedure and evaluated the effect of perceived pain on mortality and hospital length of stay (LOS). Our initial hypothesis was that increased levels of perceived pain would correlate with worse patient outcomes such as increased hospital length of stay. This would be in line with the current literature that suggest optimal pain control leads to increased mobility, earlier ambulation, and improved outcomes. [5-7] Contrary to the literature, we found that higher levels of pain were associated with reduced mortality and reduced LOS. We then discuss potential causes of these results and the general issue of dealing with counterintuitive results in retrospective data analyses.

Case

Population

We selected patients from the MIMIC database who met all of the following inclusion criteria and none of the exclusion criteria. Inclusion criteria included: (1) Adult > 18 years old, (2) who underwent CABG surgery, and (3) was extubated within 24 hours after arrival to the ICU. Exclusion criteria were: (1) Non-CABG surgical procedure, and (2) missing data on confounding variables. Patients were identified using Current Procedural Terminology (CPT) codes: The following CPT codes corresponded to the CABG procedure: 33510 to 33516 for venous grafting only for coronary artery bypass, and 33533 to 33548 for arterial grafting for coronary bypass. The final study cohort contained 844 patients (***Figure 1***).

The MIMC-III database included 1,917 patients who underwent CABG, with 844 meeting the study criteria. CABG was chosen for the investigation as it is a common procedure with the majority of patients having no or few post-operative complications and relatively predictable recoveries.[5] Due to the nature of the surgical procedure which requires sternal spreading for exposure, there is an expected high analgesic burden immediately after surgery.

Outcomes

The primary outcome assessed was mortality at 30 days. Secondary outcomes were mortality at 1

year and hospital LOS. In the MIMIC database, mortality data for patients who die after hospital discharge is derived from the social security death registry.[4]

Exposures

The exposures of interest were pain levels reported by the patient immediately and in the subsequent interval after ICU extubation. Pain levels on a scale of 0-10 were regularly self-reported by patients to ICU nurses and recorded in the database, generating a continuum of measurements for each patient. The mean, median, and maximum pain levels were used for separate analyses. Concomitant measurements of heart rates, respiratory rates, and systolic blood pressures were also compared against their simultaneously recorded pain measurement.

Intravenous (IV) opiate administration was extracted from the database. MIMIC contained data for the following medications: Morphine, fentanyl, hydromorphone, and meperidine. The was no data in MIMIC corresponding to the administration of oral analgesics.

We also looked for an association of pain and nausea for use in falsification hypothesis testing. The presence of nausea was derived from the nursing notes stored in the database. A positive nausea exposure was defined as the mention "nausea" or "nauseous" in the nursing note with no negative descriptor, such as "not nauseous" or "denies nausea", attached.

Covariates

Several variables found to be linked to worse patient outcomes in previous studies were included to control for confounding in the regression models: demographic factors, comorbid conditions, and illness severity score on admission to the ICU.[8,9] Comorbid burden was represented by the Elixhauser index which is determined by the aggregate presence or absence of 30 different comorbid conditions as detected by ICD-9 codes.[10] These conditions include but are not limited to cardiovascular disorders such as hypertension, congestive heart failure, coronary artery disease, and peripheral vascular disease; pulmonary disorders such as chronic obstructive pulmonary disease; endocrine disorders such as diabetes and hypothyroid; obesity; drug and alcohol use disorders; renal disease; liver disease. Illness severity was captured using the Oxford Acute Severity of Illness Score (OASIS), which is calculated on admission to the ICU and takes into account age, heart rate, Glasgow coma scale, mean arterial pressure, temperature, respiratory rate, ventilatory status, urine output, pre-ICU in-hospital LOS, and whether or not the patient

underwent elective surgery. Studies have shown OASIS is comparable to other illness severity ratings in predicting outcomes such as mortality and length of stay.[11]

Patient and Public Involvement

This research was done without patient or public involvement. They were not invited to contribute to the development of our methodology, our outcomes, nor the writing of our paper.

Statistical Analysis

Analysis was carried out using R version 3.4.0 and SAS 9.4. Binomial logistic regression models were fitted using maximum likelihood estimation to compare the pain measures with 30-day and 1-year mortality. Linear regression was used to model the relationship between mean pain scores and hospital LOS. Age, gender (male reference), Elixhauser index, and OASIS score were included in the models to account for potential confounders. In a separate regression, mean pain levels were categorized into four ordinal groups of no pain (0/10), mild pain (1-3), moderate pain (3-6), and severe pain (7-10) in accordance with the NIH Pain Consortium.[12]

ANOVA was used to determine if there was a significant variation in heart rate, respiratory rate, and/or systolic blood pressure, when compared to the concurrent pain assessment.

IV analgesia medications were converted to their morphine equivalents based on the National Pharmaceutical Counsel's guidelines.[13] The IV analgesia was subdivided into total dose in the first 24 hours, mean dose per ICU course day, and total dose during ICU course. ANOVA models were used to determine if there were any significant variation in administration of IV analgesics among the four categorized pain groups.

Two sensitivity analyses were performed to assess the robustness of the observed effects. The first included the same statistical tests in all postoperative CABG patients regardless of duration of intubation. The second sensitivity analysis excluded patients who died in the hospital.

To add validity to the potential observed associations, falsification hypothesis testing using Prasad and Jena's methodology was employed. A distinct and highly unlikely hypothesis is tested against the exposure of interest, pain in this case.[14] We used nausea, a symptom with no known correlation to pain, and tested it against the four different pain metrics.

Results

The database included 844 patients who underwent a CABG procedure and were extubated within 24 hours. There were 68 patients who on average reported no pain during their ICU stay after extubation, 419 with mild pain, 336 with moderate pain, and 21 with severe pain. The mean frequency of pain measurements was 19.8 measurements per patient. The distribution of patient characteristics, including age, gender, illness acuity on ICU admission (OASIS), and comorbidity index is reported in ***Table 1***. There was no significant difference noted in the frequency in which pain was assessed in those who experienced lower pain levels when compared to those who experienced increased pain levels. The number of comorbidities ranged from 0 to 9. Bivariate analysis showed increasing OASIS was significantly associated with increased mortality and increased LOS ($p < 0.05$). No significant differences were found in the amount of IV analgesia administered among the pain subgroups.

		No Pain	Mild	Moderate	Severe	p
n		68	419	336	21	
Age (mean (sd))		71.50 (10.61)	67.75 (10.54)	64.98 (9.73)	65.13 (12.85)	<0.001
Gender = male (%)		45 (66.2)	333 (79.5)	282 (83.9)	14 (66.7)	0.003
OASIS (mean (sd))		31.96 (7.25)	30.32 (6.47)	31.44 (6.35)	30.57 (6.20)	0.056
E_score (%)						<0.001
	0	4 (5.9)	96 (22.9)	87 (25.9)	7 (33.3)	
	1	12 (17.6)	116 (27.7)	97 (28.9)	4 (19.0)	
	2	12 (17.6)	81 (19.3)	79 (23.5)	4 (19.0)	
	3	10 (14.7)	61 (14.6)	46 (13.7)	3 (14.3)	
	4	12 (17.6)	29 (6.9)	16 (4.8)	1 (4.8)	
	5	6 (8.8)	19 (4.5)	8 (2.4)	2 (9.5)	
	6	7 (10.3)	8 (1.9)	2 (0.6)	0 (0.0)	
	7	2 (2.9)	4 (1.0)	1 (0.3)	0 (0.0)	
	8	0 (0.0)	4 (1.0)	0 (0.0)	0 (0.0)	
	9	3 (4.4)	1 (0.2)	0 (0.0)	0 (0.0)	
Mortality						
In Hospital (%)		9 (13.2)	5 (1.2)	1 (0.3)	0 (0.0)	<0.001
30 Day (%)		10 (14.7)	10 (2.4)	1 (0.3)	0 (0.0)	<0.001
1 Year (%)		16 (23.5)	22 (5.3)	7 (2.1)	1 (4.8)	<0.001
Narcotics						

First 24 Hrs (sd)	4.17 (5.52)	6.24 (9.85)	9.28 (25.89)	6.38 (8.07)	0.059
Daily mean (sd)	5.23 (5.43)	8.43 (7.82)	17.09 (89.87)	8.68 (8.06)	0.162
Total Narcotics (sd)	37.30 (101.39)	21.19 (70.34)	29.15 (188.08)	9.87 (8.94)	0.682

Table 1: Shows the distribution of the outcomes and covariates in the patient cohort. Abbreviations: OASIS, Oxford Acute Severity of Illness Score; E_score, Elixhauser index. OASIS score ranges from 0 to 75, with higher scores indicating more severe disease. Elixhauser index ranges from 0 to 9, with higher scores indicating a greater number of comorbid conditions.

Bivariate analysis (***Figure 2***) shows a correlation between increasing pain levels and improved outcomes among these patients who had no intra-operative complications and were extubated within 24 hours of arrival in the ICU. Higher pain levels for this specific cohort of patients who were fast-tracked after CABG were found to be associated with decreased hospital LOS. Those who experienced lower levels of pain in the ICU were more likely to be dead at 30 days and 1 year.

Multivariate regression analysis was performed to adjust for confounding. Four different models using mean, median, and maximum pain scores, and pain categories were tested against the clinical outcomes with the results displayed in ***Table 2***. The logistic regression models consistently showed that increasing pain was associated with reduced odds of death at 30 days and 1 year after adjustment for illness severity and co-morbid conditions. All the linear models demonstrated that increasing pain levels were also associated with decreased hospital LOS, except for the model that looked at the maximum pain score, which showed an opposite effect. R-Squared values for the linear regression models varied between 0.25 and 0.3 for all the models. Complete statistical data from all regression models can be found in the **online supplemental materials file.**

Model	30 Day Mortality Odds (95% Confidence Interval)	1 Year Mortality Odds (95% Confidence Interval)	Length of Stay Estimate (95% Confidence Interval)
Primary Analysis:			
Mean Pain	0.457*** (0.304 – 0.687)	0.710*** (0.571 - 0.881)	-0.916*** (-1.159, -0.673)

Median Pain	0.639*** (0.466 - 0.877)	0.856* (0.727 - 1.008)	-0.696*** (-0.886, -0.506)
Max Pain	0.812*** (0.693 - 0.951)	0.887** (0.790- 0.995)	0.148* (-0.02, 0.32)
Categorical Pain	0.214*** (0.091 - 0.502)	0.450*** (0.266 - 0.760)	-2.270*** (-2.903, 1.637)
Sensitivity Analysis 1: Including all patients regardless of intubation lengths			
Mean Pain	0.592*** (0.456 - 0.768)	0.898 (0.785 - 1.027)	-0.709*** (-0.866, -0.552)
Categorical Pain	0.328*** (0.184 - 0.586)	0.740* (0.527 - 1.037)	-1.706*** (-2.110, -1.302)
Sensitivity Analysis 2: Excluding hospital mortality patients			
Mean Pain	0.803 (0.567 - 1.137)	1.027 (0.889 - 1.187)	-0.701*** (-0.858, -0.544)
Categorical Pain	0.709 (0.309 - 1.625)	1.038 (0.714 - 1.509)	-1.680*** (-2.082, -1.278)

Table 2: Shows results from main analysis and the two sensitivity analyses.
*, **, *** denotes significance at the 90%, 95%, and 99% level, respectively.

No significant variations were noted in heart rate, respiratory rate, or blood pressure with increasing pain levels.

Sensitivity analysis was employed to examine all patients regardless of duration of intubation, expanding the sample size to 1889 patients. The results were similar for 30-day mortality and hospital LOS as regards effect size and statistical significance; however, the results were not statistically significant for 1-year mortality (***Table 2***). A total of 22 CABG patients were noted to have expired in the hospital, our cohort included 15 of these in hospital deaths. An additional sensitivity analysis excluded patients who died in the hospital- these results were consistent with the prior models and were statistically significant for hospital LOS, but not for mortality (***Table 2***).

As expected, the presence of nausea was not found to be associated with any of our pain measures in our falsification testing, decreasing the possibility that the previous results are erroneous or solely due to chance.

Discussion

Case Study

We will first discuss our unexpected results, and then discuss the general issue of counterintuitive data. Our results that increasing levels of patient-reported pain severity post-CABG surgery are associated with better clinical outcomes were not consistent with our initial hypothesis that better outcomes correlate with better pain control as per the reported literature. In fact, prior studies have found increased levels of pain in the hospital to be associated with increased mortality. [15]

The difference in the study cohort between our study and others may explain some of the discordance. Our initial analysis was limited to "fast-tracked" patients who did not have intra-operative complications and were extubated early in their ICU course. These patients made up 44% of the database patients. Studies that have reported worse clinical outcomes associated with post-operative pain did not select for a relatively healthy sub-cohort of patients. Why would patients with higher levels of pain have better outcomes? It is well documented that an increased inflammatory reaction is associated with increased pain. Pro-inflammatory cytokines such as IL-1β, IL-6, and TNF-α have been directly implicated in the physiology of pain.[16,17] These cytokines have also been found to be directly involved in wound healing through the stimulation of processes such as keratinocyte and fibroblast proliferation, and synthesis and breakdown of extracellular matrix proteins.[18] We speculate that those patients who demonstrated better outcomes mounted a more robust inflammatory response leading to more pain, but also to increased healing ability.

Another possibility is that higher perceived pain levels represent a proxy for a generally better state of health, including superior physiological function of the cardiovascular, respiratory, renal, and hepatic systems. In tandem, these systems act to metabolize and eliminate anesthetic and analgesic drugs so that the net pharmacokinetic result would likely be increased susceptibility to pain due to less administered agent remaining at active sites. Furthermore, patients with better cardiovascular function would likely have better cerebral perfusion with improved central neurological function, and thereby have a pharmacodynamic reason for perceiving more pain. Also patients who are generally in better overall condition would be expected to manifest better outcomes. These thoughts are admittedly speculative and additional research is needed to explore

these possibilities.

It is important to point out that the goal of clinicians should not be in any way to maximize pain to optimize outcomes. Conventional approaches that aim to control pain adequately should be employed. Our observation is just that - an observation of an association and conjectures of possible linking mechanisms but is not intended in any way to drive pain management policy in the direction of tolerating undertreated pain.

We performed sensitivity analyses, one including all patients regardless of post-operative ventilation duration, and another excluding patients who died during hospitalization, and reached similar conclusions. When excluding in-hospital deaths, we discovered the 30-day mortality rate had a similar odds ratio but was no longer statistically significant. This is most likely due to the low mortality rate after hospital discharge following CABG, making it difficult to detect a statistically significant effect.

We believe that researcher bias is a non-issue as these findings were not expected, but rather, the opposite. Sampling bias was also minimal. Our inclusion criteria were predefined prior to database sampling and only 28 patients needed to be excluded due to missing data. We performed multiple sensitivity analyses to determine if those that were excluded would have influenced our results. However, the study has several limitations inherent in any retrospective data analysis. We acknowledge that correlation does not equal causation and further research is needed to determine the underlying physiologic mechanism for the results seen. Due to the self-reported nature of the pain scores, reporting bias is a concern. Some patients may have over-reported and others under-reported their pain. We also recognize that analgesic administration is a confounder and were unable to completely control for this due to lack of information regarding oral analgesics in the database However, with respect to intravenous analgesics, we attempted to limit this potential confounder by excluding those with prolonged intubations who would inherently have received and required greater doses of sedatives and analgesics. We also compared the amount of narcotics that patients were receiving and did not observe any significant differences among the various pain groups. Despite measures taken to guarantee internal validity, we anticipate appropriate skepticism with regard to generalizability of the findings. This, of course, is of genuine concern given the current state-of-affairs where clinicians are already inundated with conflicting studies of questionable quality. We therefore invite other

investigators to replicate (and expand) our analysis in other databases.

Counterintuitive Results and examples

As noted, our findings were contrary to clinical expectations and to most published works which associate increased pain with worse outcomes.[15,19-20] Encountering counterintuitive results is not unique to retrospective data analysis. Clinicians encounter unexpected, possibly aberrant, values in situations such as the evaluation of laboratory and monitor data. When a possibly spurious lab result is obtained, the usual response is to repeat the test. When the second test comes back with a more acceptable value, we generally then ignore the unexpected value. But what if the repeat value is also aberrant? Do we repeat it again, or do we begin to believe that the value is 'real' and start to formulate a response to a clinical problem? In this case, it is the *consistency* and *reproducibility* of the counterintuitive value that drives its possible validity. The details of this process are determined by the overall clinical risks involved. The consistency we found in the pain score values drove us to consider the possibility that the values were 'real' even though they were counterintuitive in terms of our expectations.

Another issue in evaluating to counterintuitive values is whether they are *possible.* Impossible values would include a potassium of 64.5, one incompatible with life. But a potassium of 7.3 is a possible value. The pain values associated with better outcomes were unexpected, but not so high that they were impossible in the observed context.

One question that would arise with a potassium of 7.3 would be that of continuity- did the value occur suddenly or gradually in a stream of normal values? Were surrounding values similarly abnormal? In the context of persistently abnormal values, e.g. untreated uremia, a normal value would be counterintuitive. So that while most counterintuitive values will tend to be out of the 'normal range', they will not necessarily be so. In the context of increasing values, it might simply be the first one that was not only out of the normal range, but that crossed the line into a critical range,

The fundamental question is whether counterintuitive results are actually false results, or does the problem lie in our perception of what should be. **Table 3** displays a categorization of error types that could result in faulty data. We are not able to attribute the counterintuitive data we observed to any of these factors, however.

Human error	Mis-entry; misunderstanding of scale values; faulty understanding of use of data entry software; faulty interpretation of device values
Lab error	Sampling error (e.g. hemolysis); measurement error
Device error	Disconnect, interference, faulty calibration, software error; unexplained, transient aberrant values that resolve and do not recur
Systems error	Interface error, application interoperability error
Software error	Bug in software relating to data value entry; data wrongly captured, stored, and/or retrieved due to software design faults or bugs
Hardware error	Hardware issues that impact software and systems
Data analytic error	Error in analytic algorithm or process

Table 3: Putative causes of truly faulty data

How can counterintuitive results be approached in secondary data analyses? **Table 4** displays characteristics that may distinguish reliable (but counterintuitive) from truly faulty data. With consideration of these factors, the first investigative step is to retrace the process and workflow involved in data entry so far as possible. Our data was obtained at the institution of several of the authors where nurses are trained to assess pain on a standard scale from 0 to 10. There are several potential faults to this method. The nursing staff could neglect to regularly assess pain or neglect to enter the information into the medical record generating the database. While this may alter a few data points, it is unlikely to systematically affect all data unless there was an obvious glaring institutional issue affecting every nurse and every data entry.

Viability	Is the value consistent with clinical reality? Are the values even possible ones?
Consistency	If applicable (not always the case in retrospective analysis), is the value observed

	consistently, such as in our pain score observations?
Continuity	What is the context of the value- does it occur as a sudden aberrant value (a 'blip'), or as one of increasingly aberrant values (a trend)?
Identity	Are the circumstances that produced the data truly identical so far as identifiable? I.e. Would the same circumstances produce the same data results in a different database, institutional, or cultural context?
Reproducibility	Is the value reproducible on repetition? while reproduction cannot be performed upon retrospective data, can the values be reproduced upon observation across different clinical databases, or in the same database over ongoing time?
Sensibility	Even if it does not meet current clinical expectations, does it make potential sense in associated clinical context?
Curiosity	Does it drive the observer to seek alternative better solutions and pose questions for further research?

Table 4: Criteria to establish possible validity of counterintuitive data

After determining that the data source is valid, additional statistical tests can be run on the patient cohort. Tests such as the falsification hypothesis testing we utilized, add validity to the results as they show that the cohort follows other generally known principles. In our study, falsification analysis provided support for our findings.

Concurrent contextual data can also help to confirm the veracity of data- for example, one could examine ECGs if hyperkalemia was being analyzed. We examined concomitant vital signs during the time of pain measurements. We expected to observe significant increases with higher pain levels, but did not: With the combination of analgesics, residual anesthetics, and the concurrent use of drugs that directly affect vital signs such as beta-blockers, the lack of correlation is probably not surprising. In fact, we learned that in this setting, it appears to be

inadvisable to use vital sign changes as a proxy for the presence of unvoiced pain. Finally, one can attempt to physiologically explain the disparity between the observed and expected results as we did above for the case of post- CABG pain.

The use of lower thresholds for blood transfusions in the ICU is an example of a counterintuitive finding. ICU target hemoglobin levels were historically set at greater than 10 g/dL, theoretically to ensure adequate oxygen delivery.[21] This led to increased transmission of blood borne diseases, unnecessary healthcare expenditures, and actually worse outcomes.[22] Later research showed that this rule was not necessary for most patients, but only for selected patients such as those with acute coronary syndrome actively experiencing chest pain. The initially counterintuitive findings that lower hemoglobin levels were not only acceptable but preferable in most cases, served as research triggers to more fully elucidate optimal clinical practice. Our case may serve as an analogous research trigger in terms of optimally managing postoperative pain. Outcomes such as mortality and LOS are complex phenomena driven by many factors- to observe a clear and robust statistical effect such as we did is strongly suggestive that something 'real' is occurring even if the data were counterintuitive.

The final step when dealing with counterintuitive data is to look for additional evidence that confirms the reliability of the results (perhaps this could be termed 'confirmatory metadata'). With respect to our CABG case, the analysis should be rerun on additional databases and in different settings. Just as clinicians continued to manage intensive care unit anemia as they always had until more definitive results were reported, our results should not impact the analgesic care of patients at this point. However, we hope that we have raised the issue in the appropriate minds that outcomes may benefit from approaches slightly different from usual. After all, one can easily eliminate all pain from postoperative patients but they would have to remain sedated and ventilated for an indefinite period of time to do so. And after they are extubated, pain management should not be so aggressive that it leads to apnea and respiratory arrest. In other words, there may be a detectable level of tolerable pain that leads patients to their best outcomes, and no honest clinician will guarantee a patient that they will have no pain at all after a procedure like a sternal-disrupting CABG.

Conclusion

Contrary to our expectations, we observed, in a retrospective analysis of electronic health records, that post-CABG fast-track patients with higher pain scores had better outcomes. The increasing use of EHRs for secondary analysis will likely lead to an increasing incidence of such apparently counterintuitive results. While the first step in this situation is to attempt to confirm the reliability of both the analytic process and the data itself, such findings that prove to be robust may lead to further ideas and subsequent research that drive future clinical care. On the other hand, clinicians must be careful in terms of modifying their practices until the implications of such counterintuitive (or any) data have been thoroughly vetted and confirmed in diverse database contexts and via the peer review process.

Declarations

Availability of Data Materials

The datasets generated for the current study were derived from the MIMIC-III Database available at https://mimic.physionet.org/. The data subsets and statistical code used in this project can be found at https://github.com/ErikWDoty/PainProject.

Consent for Publication

Not applicable

Competing Interests

The authors declare they have no completing interests.

Funding

There was no financial support for this project.

Author Contributions

ED was responsible for the data extraction, the initial statistical analysis, and writing and editing the manuscript. NM was involved in validating the statistical models and participated in editing the manuscript. DS was responsible for assisting with background information and editing the manuscript. LC was the project supervisor, responsible for project conception and manuscript editing.

Acknowledgements

We would like to thank J. Michael Jaeger, MD, PhD of the University of Virginia School of Medicine, for his assistance with background information.

References

[1] Gajic O, Dara SI, Mendez JL, et al. Ventilator-associated lung injury in patients without acute lung injury at the onset of mechanical ventilation. Critical Care Medicine. 2004;32:1817-1824.

[2] A.F Connors Jr, T Speroff, N.V Dawson, et al. The effectiveness of right heart catheterization in the initial care of critically ill patients (SUPPORT Investigators) JAMA, 276 (1996), pp. 889-89

[3] Hine LK, Laird N, Hewitt P, Chalmers TC. Meta-analytic Evidence Against Prophylactic Use of Lidocaine in Acute Myocardial Infarction. Arch Intern Med. 1989;149(12):2694–2698. doi:10.1001/archinte.1989.00390120056011

[4] MIMIC-III, a freely accessible critical care database. Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. Scientific Data (2016). DOI: 10.1038/sdata.2016.35. Available from: http://www.nature.com/articles/sdata201635

[5] Sutherland, J., Liu, G., Crump, T., Bair, M., & Karimuddin, A. (2018). Relationship between preoperative patient-reported outcomes and hospital length of stay: A prospective cohort study of general surgery patients in Vancouver, Canada. *Journal of Health Services Research & Policy*. doi:10.1177/1355819618791634

[6] Bykowski, M. R., Sivak, W., Garland, C., Cladis, F. P., Goldstein, J. A., & Losee, J. E. (2018). A Multimodal Preemptive Analgesic Protocol for Alveolar Bone Graft Surgery: Decreased Pain, Hospital Stay, and Health Care Costs. *The Cleft Palate-Craniofacial Journal*. doi:10.1177/1055665618791943

[7] Ljungqvist O, Scott M, Fearon KC. (2017) Enhanced Recovery After Surgery: A

Review. *JAMA Surg*;152(3):292–298. doi:10.1001/jamasurg.2016.4952

[8] Hawkes, A. L., Nowak, M., Bidstrup, B., & Speare, R. (2006). Outcomes of coronary artery bypass graft surgery. Vascular Health and Risk Management, 2(4), 477–484.Hawkes, A. L., Nowak, M., Bidstrup, B., & Speare, R. (2006). Outcomes of coronary artery bypass graft surgery. *Vascular Health and Risk Management*, *2*(4), 477–484.

[9] Shroyer, A. W., Coombs, L. P., Peterson, E. D., Eiken, M. C., Delong, E. R., Chen, A., Edwards, F. H. (2003). The society of thoracic surgeons: 30-day operative mortality and morbidity risk models. The Annals of Thoracic Surgery, 75(6), 1856-1865. doi:10.1016/s0003-4975(03)00179-6

[10] Elixhauser, A., Steiner, C., Harris, D., & Coffey, R. (1998). Comorbidity Measures for Use with Administrative Data. Medical Care, 36(1), 8-27. Retrieved from http://www.jstor.org/stable/3766985

[11] Johnson, A. E., Kramer, A. A., & Clifford, G. D. (2013). A New Severity of Illness Scale Using a Subset of Acute Physiology and Chronic Health Evaluation Data Elements Shows Comparable Predictive Accuracy*. Critical Care Medicine, 41(7), 1711-1718. doi:10.1097/ccm.0b013e31828a24fe

[12] Pain Intensity Instruments. (2003, July). Retrieved April 9, 2017, from https://painconsortium.nih.gov/pain_scales/NumericRatingScale.pdf

[13] *Pain: Current understanding of assessment, management, and treatments*. (2001). Reston, VA: National Pharmaceutical Council.

[14] Prasad, V., & Jena, A. B. (2013). Prespecified Falsification End Points. *Jama, 309*(3), 241. doi:10.1001/jama.2012.96867

[15] Lewis KS, Whipple JK, Michael KA, Quebbeman EJ. Effect of analgesic treatment on

the physiological consequences of acute pain. Am J Hosp Pharm. 1994;51(12):1539-54.

[16] Watkins LR, Milligan ED, Maier SF. (2003). Glial proinflammatory cytokines mediate exaggerated pain states: implications for clinical pain. *Advances in Experimental Medicine and Biology*. 521: 1-21

[17] Zhang, J.-M., & An, J. (2007). Cytokines, Inflammation and Pain. *International Anesthesiology Clinics*, *45*(2), 27–37. http://doi.org/10.1097/AIA.0b013e318034194e

[18] Werner, S., & Grose, R. (2003). Regulation of Wound Healing by Growth Factors and Cytokines. *American Physiological Society, 83*(3), 835-870.

[19] Rodgers, A. (2000). Reduction of postoperative mortality and morbidity with epidural or spinal anaesthesia: results from overview of randomised. Bmj, 321(7275), 1493-1493. 2doi:10.1136/bmj.321.7275.1493

[20] Wu, C., Hurley, R., Anderson, G., Herbert, R., Rowlingson, A., & Fleisher, L. (2004). Effect of postoperative epidural analgesia on morbidity and mortality following surgery in medicare patients. Regional Anesthesia and Pain Medicine, 29(6), 525-533. doi:10.1016/j.rapm.2004.07.002

[21] Wang J, Klein H. Red blood cell transfusion in the treatment and management of anaemia: the search for the elusive transfusion trigger. Vox Sanguinis [serial online]. January 2010;98(1):2-11. Available from: Academic Search Complete, Ipswich, MA. Accessed June 8, 2018.

[22] Seitz, K. P., Sevransky, J. E., Martin, G. S., Roback, J. D., & Murphy, D. J. (2017). Evaluation of RBC Transfusion Practice in Adult ICUs and the Effect of Restrictive Transfusion Protocols on Routine Care. *Critical Care Medicine*, *45*(2), 271–281. http://doi.org/10.1097/CCM.0000000000002077

Captions

Figure 1: Shows selection of patient cohort from MIMIC Database. After selecting those who underwent CABG procedure and excluding those with no pain measurements; 844 patients were extubated within 24 hours following surgery and included in the cohort.

Figure 2: Three plots demonstrating the bivariate relationship between the outcomes of interest and mean pain. Plot A shows decreased length of stays with increased mean pain levels. Plot B and Plot C show that, on average, those who expired at 30 days and 1 year marks experienced lower in hospital pain levels than those who did not expire.

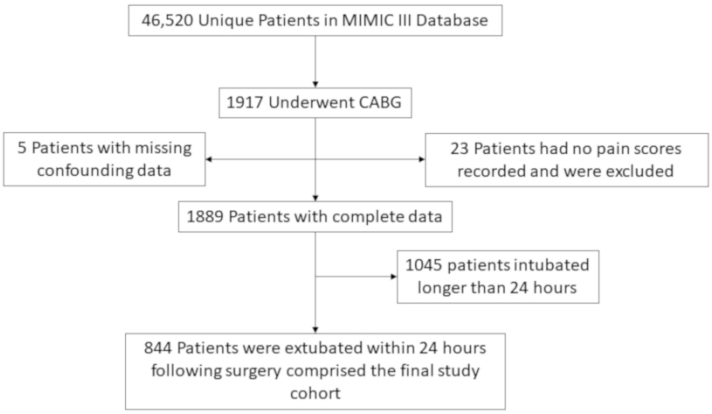
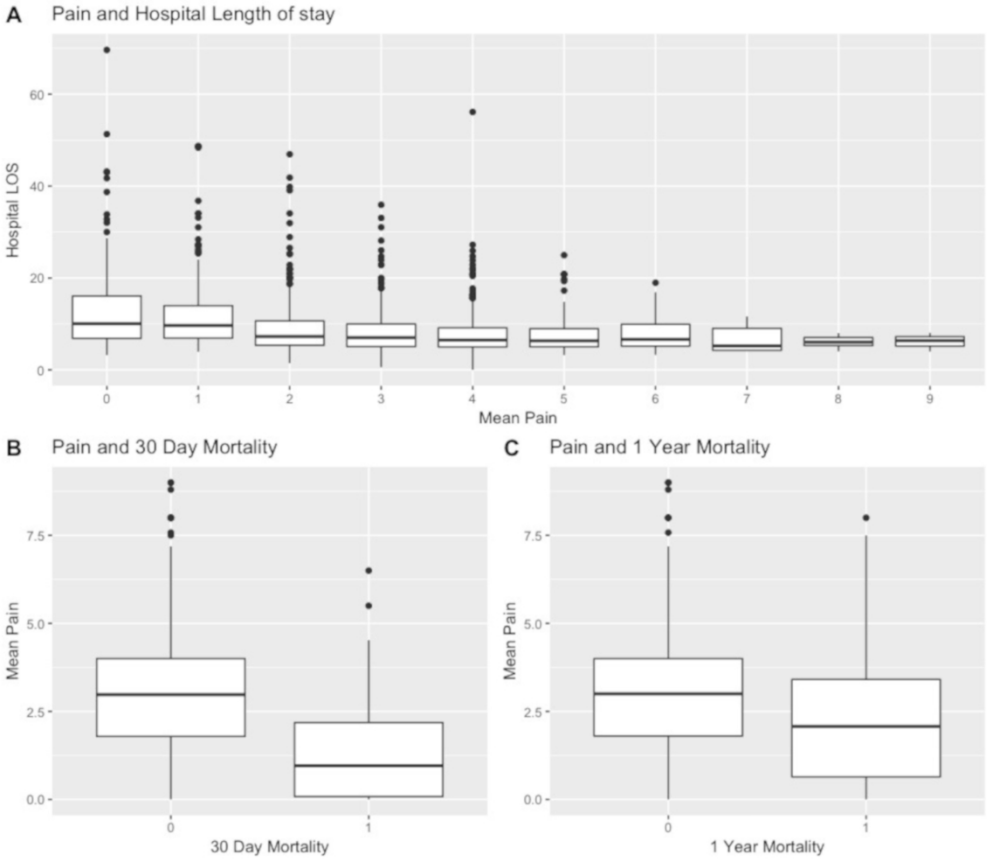


Figure 1: Shows the selection of our study cohort and those eliminated for missing data

114x64mm (600 x 600 DPI)



Three plots demonstrating the bivariate relationship between the outcomes of interest and mean pain. Plot A shows decreased length of stays with increased mean pain levels. Plot B and Plot C show that, on average, those who expired at 30 days and 1 year marks experienced lower in hospital pain levels than those who did not expire.

114x99mm (600 x 600 DPI)

Counterintuitive Results From Observational Data: A Case Study and Discussion – Online Supplemental File

Summary: The following is a complete statistical output from the data above. This file includes the primary model results including the multiple regression models comparing mean, median, maximum and the categorical pain levels to the studied outcomes, included mortality and length of stay. Also included are the results from the sensitivity analysis in which all CABG patients were included and patients who expired in the hospital were excluded.

Table of Contents

Model 1: Mean pain vs Hospital LOS 2
Model 2: Median Pain vs Hospital LOS 3
Model 3: Maximum pain vs Hospital LOS 4
Model 4: Categorical Pain vs Hospital LOS 5
Model 5: Mean pain vs 30-day mortality 6
Model 6: Median Pain vs 30-day Mortality 8
Model 7: Maximum pain vs 30-day Mortality 10
Model 8: Categorical Pain vs 30-day Mortality 12
Model 9: Mean Pain vs 1-yr Mortality 14
Model 10: Median Pain vs 1-yr Mortality 16
Model 11: Maximum Pain vs 1-yr Mortality 18
Model 12: Categorical Pain vs 1-yr Mortality 20
Sensitivity Model 1: Mean pain vs Hospital LOS 22
Sensitivity Model 2: Mean Pain vs 30-day Mortality 23
Sensitivity Model 3: Mean Pain vs 1-yr Mortality 26
Sensitivity Model 4: Categorical Pain vs Hospital LOS 28
Sensitivity Model 5: Categorical Pain vs 30-day Mortality 29
Sensitivity Model 6: Categorical pain vs 1-yr Mortality 31
Sensitivity Model 7: Mean pain vs Hospital Length of Stay 33
Sensitivity Model 8: Mean Pain vs 30-day Mortality 34
Sensitivity Model 9: Mean Pain vs 1-yr Mortality 36
Sensitivity Model 10: Categorical Pain vs Hospital LOS 38
Sensitivity Model 11: Categorical Pain vs 30-day Mortality 39
Sensitivity Model 12: Categorical Pain vs 1-yr Mortality 41

Model 1: Mean pain vs Hospital LOS

Number of Observations Read	844
Number of Observations Used	844

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	11888	2377.65640	70.77	<.0001
Error	838	28155	33.59817		
Corrected Total	843	40044			

Root MSE	5.79639	R-Square	0.2969
Dependent Mean	8.58776	Adj R-Sq	0.2927
Coeff Var	67.49599		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	6.74721	1.70464	3.96	<.0001
mean_pain	1	-0.91633	0.12415	-7.38	<.0001
male	1	-1.78286	0.51402	-3.47	0.0006
age	1	0.00471	0.02021	0.23	0.8160
e_score	1	1.61599	0.12331	13.10	<.0001
oasis	1	0.09119	0.03159	2.89	0.0040

Model 2: Median Pain vs Hospital LOS

Number of Observations Read	844
Number of Observations Used	844

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	11808	2361.65507	70.09	<.0001
Error	838	28235	33.69364		
Corrected Total	843	40044			

Root MSE	5.80462	R-Square	0.2949
Dependent Mean	8.58776	Adj R-Sq	0.2907
Coeff Var	67.59182		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	5.51742	1.66209	3.32	0.0009
med_pain	1	-0.69605	0.09657	-7.21	<.0001
male	1	-1.75771	0.51489	-3.41	0.0007
age	1	0.01249	0.02011	0.62	0.5346
e_score	1	1.62356	0.12339	13.16	<.0001
oasis	1	0.08689	0.03159	2.75	0.0061

Model 3: Maximum pain vs Hospital LOS

Number of Observations Read	844
Number of Observations Used	844

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	10158	2031.60013	56.97	<.0001
Error	838	29886	35.66294		
Corrected Total	843	40044			

Root MSE	5.97185	R-Square	0.2537
Dependent Mean	8.58776	Adj R-Sq	0.2492
Coeff Var	69.53905		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.28286	1.84430	0.70	0.4869
max_pain	1	0.14781	0.08819	1.68	0.0941
male	1	-1.91709	0.52929	-3.62	0.0003
age	1	0.03550	0.02098	1.69	0.0910
e_score	1	1.79329	0.12447	14.41	<.0001
oasis	1	0.06871	0.03250	2.11	0.0348

Model 4: Categorical Pain vs Hospital LOS

Number of Observations Read	844
Number of Observations Used	844

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	11727	2345.48240	69.41	<.0001
Error	838	28316	33.79014		
Corrected Total	843	40044			

Root MSE	5.81293	R-Square	0.2929
Dependent Mean	8.58776	Adj R-Sq	0.2886
Coeff Var	67.68854		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	6.47649	1.70556	3.80	0.0002
cat_pain	1	-2.26964	0.32289	-7.03	<.0001
male	1	-1.78270	0.51551	-3.46	0.0006
age	1	0.00679	0.02025	0.34	0.7376
e_score	1	1.62244	0.12372	13.11	<.0001
oasis	1	0.09063	0.03168	2.86	0.0043

Model 5: Mean pain vs 30-day mortality

Number of Observations Read	844
Number of Observations Used	844

Response Profile		
Ordered Value	X30_day	Total Frequency
1	1	21
2	0	823

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	198.606	149.641
SC	203.344	178.070
-2 Log L	196.606	137.641

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	58.9644	5	<.0001
Score	72.0933	5	<.0001
Wald	40.8033	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-7.3196	2.3007	10.1217	0.0015
mean_pain	1	-0.7830	0.2078	14.1967	0.0002
age	1	0.0268	0.0255	1.0995	0.2944
male	1	0.5256	0.5659	0.8627	0.3530
e_score	1	0.4041	0.1115	13.1357	0.0003
oasis	1	0.0553	0.0356	2.4204	0.1198

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
mean_pain	0.457	0.304	0.687
age	1.027	0.977	1.080
male	1.692	0.558	5.128
e_score	1.498	1.204	1.864
oasis	1.057	0.986	1.133

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	91.1	Somers' D	0.821
Percent Discordant	8.9	Gamma	0.821
Percent Tied	0.0	Tau-a	0.040
Pairs	17283	c	0.911

Partition for the Hosmer and Lemeshow Test					
		X30_day = 1		X30_day = 0	
Group	Total	Observed	Expected	Observed	Expected
1	84	0	0.04	84	83.96
2	84	0	0.09	84	83.91
3	84	0	0.15	84	83.85
4	84	0	0.24	84	83.76
5	84	1	0.36	83	83.64
6	84	0	0.54	84	83.46
7	84	0	0.84	84	83.16
8	84	3	1.45	81	82.55
9	84	3	3.13	81	80.87
10	88	14	14.16	74	73.84

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
4.7470	8	0.7842

Model 6: Median Pain vs 30-day Mortality

Number of Observations Read	844
Number of Observations Used	844

Response Profile		
Ordered Value	X30_day	Total Frequency
1	1	21
2	0	823

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	198.606	158.742
SC	203.344	187.171
-2 Log L	196.606	146.742

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	49.8631	5	<.0001
Score	63.0052	5	<.0001
Wald	38.4675	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-8.9316	2.1971	16.5261	<.0001
med_pain	1	-0.4474	0.1612	7.7053	0.0055
age	1	0.0377	0.0253	2.2238	0.1359
male	1	0.5085	0.5589	0.8276	0.3630
e_score	1	0.4428	0.1083	16.7092	<.0001
oasis	1	0.0548	0.0348	2.4831	0.1151

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
med_pain	0.639	0.466	0.877
age	1.038	0.988	1.091
male	1.663	0.556	4.973
e_score	1.557	1.259	1.925
oasis	1.056	0.987	1.131

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	88.4	Somers' D	0.768
Percent Discordant	11.6	Gamma	0.768
Percent Tied	0.0	Tau-a	0.037
Pairs	17283	c	0.884

Partition for the Hosmer and Lemeshow Test					
		X30_day = 1		X30_day = 0	
Group	Total	Observed	Expected	Observed	Expected
1	84	0	0.07	84	83.93
2	84	0	0.15	84	83.85
3	84	0	0.23	84	83.77
4	84	0	0.33	84	83.67
5	84	1	0.49	83	83.51
6	84	0	0.69	84	83.31
7	84	3	1.05	81	82.95
8	84	1	1.70	83	82.30
9	84	3	3.37	81	80.63
10	88	13	12.91	75	75.09

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
6.0151	8	0.6455

Model 7: Maximum pain vs 30-day Mortality

Number of Observations Read	844
Number of Observations Used	844

Response Profile		
Ordered Value	X30_day	Total Frequency
1	1	21
2	0	823

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	198.606	162.636
SC	203.344	191.065
-2 Log L	196.606	150.636

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	45.9693	5	<.0001
Score	61.5400	5	<.0001
Wald	39.1926	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-8.0597	2.2457	12.8804	0.0003
max_pain	1	-0.2081	0.0806	6.6669	0.0098
age	1	0.0276	0.0267	1.0721	0.3005
male	1	0.2187	0.5421	0.1628	0.6866
e_score	1	0.5779	0.1088	28.2346	<.0001
oasis	1	0.0587	0.0344	2.9173	0.0876

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
max_pain	0.812	0.693	0.951
age	1.028	0.976	1.083
male	1.244	0.430	3.601
e_score	1.782	1.440	2.206
oasis	1.060	0.991	1.134

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	86.8	Somers' D	0.736
Percent Discordant	13.2	Gamma	0.736
Percent Tied	0.0	Tau-a	0.036
Pairs	17283	c	0.868

Partition for the Hosmer and Lemeshow Test					
		X30_day = 1		X30_day = 0	
Group	Total	Observed	Expected	Observed	Expected
1	84	0	0.14	84	83.86
2	84	0	0.24	84	83.76
3	84	1	0.36	83	83.64
4	84	1	0.47	83	83.53
5	84	0	0.62	84	83.38
6	84	0	0.84	84	83.16
7	84	0	1.17	84	82.83
8	84	2	1.75	82	82.25
9	84	4	3.21	80	80.79
10	88	13	12.20	75	75.80

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
5.0983	8	0.7470

Model 8: Categorical Pain vs 30-day Mortality

Number of Observations Read	844
Number of Observations Used	844

Response Profile		
Ordered Value	X30_day	Total Frequency
1	1	21
2	0	823

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	198.606	154.226
SC	203.344	182.654
-2 Log L	196.606	142.226

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	54.3800	5	<.0001
Score	69.7089	5	<.0001
Wald	42.5015	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-7.9103	2.2412	12.4572	0.0004
cat_pain	1	-1.5417	0.4355	12.5328	0.0004
age	1	0.0298	0.0255	1.3654	0.2426
male	1	0.5105	0.5639	0.8196	0.3653
e_score	1	0.4277	0.1127	14.4131	0.0001
oasis	1	0.0560	0.0345	2.6370	0.1044

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
cat_pain	0.214	0.091	0.502
age	1.030	0.980	1.083
male	1.666	0.552	5.031
e_score	1.534	1.230	1.913
oasis	1.058	0.988	1.132

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	90.7	Somers' D	0.814
Percent Discordant	9.3	Gamma	0.814
Percent Tied	0.0	Tau-a	0.040
Pairs	17283	c	0.907

Partition for the Hosmer and Lemeshow Test					
		X30_day = 1		X30_day = 0	
Group	Total	Observed	Expected	Observed	Expected
1	84	0	0.07	84	83.93
2	85	0	0.14	85	84.86
3	84	0	0.23	84	83.77
4	84	0	0.37	84	83.63
5	84	0	0.52	84	83.48
6	84	1	0.69	83	83.31
7	84	0	1.02	84	82.98
8	84	2	1.51	82	82.49
9	84	3	2.93	81	81.07
10	87	15	13.52	72	73.48

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
2.8512	8	0.9433

Model 9: Mean Pain vs 1-yr Mortality

Number of Observations Read	844
Number of Observations Used	844

Response Profile		
Ordered Value	X1_yr	Total Frequency
1	1	46
2	0	798

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	359.121	282.785
SC	363.859	311.214
-2 Log L	357.121	270.785

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	86.3361	5	<.0001
Score	100.3926	5	<.0001
Wald	64.9324	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-9.2928	1.6644	31.1714	<.0001
mean_pain	1	-0.3430	0.1105	9.6411	0.0019
age	1	0.0599	0.0191	9.8411	0.0017
male	1	0.3160	0.3883	0.6622	0.4158
e_score	1	0.4610	0.0834	30.5477	<.0001
oasis	1	0.0496	0.0243	4.1861	0.0408

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
mean_pain	0.710	0.571	0.881
age	1.062	1.023	1.102
male	1.372	0.641	2.936
e_score	1.586	1.347	1.867
oasis	1.051	1.002	1.102

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	85.2	Somers' D	0.704
Percent Discordant	14.8	Gamma	0.704
Percent Tied	0.0	Tau-a	0.073
Pairs	36708	c	0.852

Partition for the Hosmer and Lemeshow Test					
		X1_yr = 1		X1_yr = 0	
Group	Total	Observed	Expected	Observed	Expected
1	84	0	0.23	84	83.77
2	84	0	0.49	84	83.51
3	84	0	0.78	84	83.22
4	84	3	1.11	81	82.89
5	84	2	1.50	82	82.50
6	84	1	2.04	83	81.96
7	84	3	2.94	81	81.06
8	84	3	4.67	81	79.33
9	84	10	7.86	74	76.14
10	88	24	24.36	64	63.64

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
6.7640	8	0.5623

Model 10: Median Pain vs 1-yr Mortality

Number of Observations Read	844
Number of Observations Used	844

Response Profile		
Ordered Value	X1_yr	Total Frequency
1	1	46
2	0	798

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	359.121	289.827
SC	363.859	318.256
-2 Log L	357.121	277.827

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	79.2945	5	<.0001
Score	92.8046	5	<.0001
Wald	62.5204	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-10.0363	1.6198	38.3892	<.0001
med_pain	1	-0.1552	0.0835	3.4581	0.0629
age	1	0.0647	0.0190	11.5703	0.0007
male	1	0.2607	0.3844	0.4600	0.4976
e_score	1	0.4868	0.0828	34.5309	<.0001
oasis	1	0.0461	0.0239	3.7018	0.0544

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
med_pain	0.856	0.727	1.008
age	1.067	1.028	1.107
male	1.298	0.611	2.757
e_score	1.627	1.383	1.914
oasis	1.047	0.999	1.097

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	83.6	Somers' D	0.672
Percent Discordant	16.4	Gamma	0.672
Percent Tied	0.0	Tau-a	0.069
Pairs	36708	c	0.836

Partition for the Hosmer and Lemeshow Test					
		X1_yr = 1		X1_yr = 0	
Group	Total	Observed	Expected	Observed	Expected
1	84	0	0.28	84	83.72
2	84	1	0.60	83	83.40
3	84	0	0.89	84	83.11
4	84	2	1.23	82	82.77
5	84	3	1.63	81	82.37
6	84	1	2.24	83	81.76
7	84	3	3.24	81	80.76
8	84	2	4.90	82	79.10
9	84	10	7.45	74	76.55
10	88	24	23.55	64	64.45

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
6.6165	8	0.5785

Model 11: Maximum Pain vs 1-yr Mortality

Number of Observations Read	844
Number of Observations Used	844

Response Profile		
Ordered Value	X1_yr	Total Frequency
1	1	46
2	0	798

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	359.121	289.427
SC	363.859	317.856
-2 Log L	357.121	277.427

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	79.6942	5	<.0001
Score	93.2270	5	<.0001
Wald	62.4345	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-9.4442	1.6770	31.7153	<.0001
max_pain	1	-0.1205	0.0591	4.1619	0.0413
age	1	0.0596	0.0194	9.3917	0.0022
male	1	0.1720	0.3776	0.2074	0.6488
e_score	1	0.5437	0.0822	43.7465	<.0001
oasis	1	0.0487	0.0241	4.0913	0.0431

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
max_pain	0.887	0.790	0.995
age	1.061	1.022	1.103
male	1.188	0.567	2.490
e_score	1.722	1.466	2.023
oasis	1.050	1.002	1.101

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	83.1	Somers' D	0.663
Percent Discordant	16.9	Gamma	0.663
Percent Tied	0.0	Tau-a	0.068
Pairs	36708	c	0.831

Partition for the Hosmer and Lemeshow Test					
		X1_yr = 1		X1_yr = 0	
Group	Total	Observed	Expected	Observed	Expected
1	84	1	0.29	83	83.71
2	84	0	0.58	84	83.42
3	85	2	0.90	83	84.10
4	84	1	1.23	83	82.77
5	84	0	1.68	84	82.32
6	84	4	2.26	80	81.74
7	84	2	3.16	82	80.84
8	84	4	4.75	80	79.25
9	84	8	7.94	76	76.06
10	87	24	23.21	63	63.79

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
7.4004	8	0.4941

Model 12: Categorical Pain vs 1-yr Mortality

Number of Observations Read	844
Number of Observations Used	844

Response Profile		
Ordered Value	X1_yr	Total Frequency
1	1	46
2	0	798

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	359.121	284.013
SC	363.859	312.442
-2 Log L	357.121	272.013

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	85.1076	5	<.0001
Score	99.4422	5	<.0001
Wald	64.6025	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-9.4525	1.6507	32.7923	<.0001
cat_pain	1	-0.7994	0.2680	8.8971	0.0029
age	1	0.0605	0.0190	10.1350	0.0015
male	1	0.3156	0.3878	0.6624	0.4157
e_score	1	0.4689	0.0836	31.4847	<.0001
oasis	1	0.0501	0.0241	4.3314	0.0374

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
cat_pain	0.450	0.266	0.760
age	1.062	1.024	1.103
male	1.371	0.641	2.932
e_score	1.598	1.357	1.883
oasis	1.051	1.003	1.102

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	85.1	Somers' D	0.702
Percent Discordant	14.9	Gamma	0.702
Percent Tied	0.0	Tau-a	0.072
Pairs	36708	c	0.851

Partition for the Hosmer and Lemeshow Test					
		X1_yr = 1		X1_yr = 0	
Group	Total	Observed	Expected	Observed	Expected
1	84	0	0.25	84	83.75
2	84	0	0.52	84	83.48
3	84	0	0.78	84	83.22
4	84	4	1.12	80	82.88
5	84	1	1.55	83	82.45
6	84	1	2.11	83	81.89
7	84	1	3.14	83	80.86
8	84	5	4.69	79	79.31
9	84	10	7.81	74	76.19
10	88	24	24.03	64	63.97

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
12.1299	8	0.1455

Sensitivity Model 1: Mean pain vs Hospital LOS

All CABG patients included

Number of Observations Read	1889
Number of Observations Used	1889

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	13989	2797.85361	96.07	<.0001
Error	1883	54838	29.12252		
Corrected Total	1888	68827			

Root MSE	5.39653	R-Square	0.2033
Dependent Mean	9.05966	Adj R-Sq	0.2011
Coeff Var	59.56654		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	6.13876	1.07024	5.74	<.0001
mean_pain	1	-0.70945	0.08044	-8.82	<.0001
male	1	-1.08190	0.29793	-3.63	0.0003
age	1	0.01732	0.01270	1.36	0.1728
e_score	1	1.13959	0.07364	15.47	<.0001
oasis	1	0.07134	0.01895	3.76	0.0002

Sensitivity Model 2: Mean Pain vs 30-day Mortality

All CABG patient included

Number of Observations Read	1889
Number of Observations Used	1889

Response Profile

Ordered Value	X30_day	Total Frequency
1	1	38
2	0	1851

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	374.103	324.079
SC	379.647	357.342
-2 Log L	372.103	312.079

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	60.0235	5	<.0001
Score	66.7408	5	<.0001
Wald	53.0781	5	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-6.5574	1.5435	18.0499	<.0001
mean_pain	1	-0.5241	0.1328	15.5838	<.0001
age	1	0.0188	0.0180	1.0847	0.2977
male	1	-0.0844	0.3616	0.0545	0.8154
e_score	1	0.3246	0.0785	17.0945	<.0001
oasis	1	0.0482	0.0244	3.9116	0.0480

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
mean_pain	0.592	0.456	0.768
age	1.019	0.984	1.056
male	0.919	0.452	1.867
e_score	1.384	1.186	1.614
oasis	1.049	1.000	1.101

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	81.1	Somers' D	0.622
Percent Discordant	18.9	Gamma	0.622
Percent Tied	0.0	Tau-a	0.025
Pairs	70338	c	0.811

Partition for the Hosmer and Lemeshow Test					
		X30_day = 1		X30_day = 0	
Group	Total	Observed	Expected	Observed	Expected
1	190	1	0.26	189	189.74
2	189	0	0.51	189	188.49
3	189	0	0.76	189	188.24
4	189	2	1.01	187	187.99
5	189	1	1.39	188	187.61
6	189	0	1.90	189	187.10
7	189	8	2.71	181	186.29
8	189	2	4.04	187	184.96
9	189	4	6.87	185	182.13
10	187	20	18.56	167	168.44

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
19.2358	8	0.0136

Sensitivity Model 3: Mean Pain vs 1-yr Mortality

All CABG patients included

Number of Observations Read	1889
Number of Observations Used	1889

Response Profile		
Ordered Value	X1_yr	Total Frequency
1	1	104
2	0	1785

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	807.244	702.254
SC	812.788	735.517
-2 Log L	805.244	690.254

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	114.9898	5	<.0001
Score	129.4134	5	<.0001
Wald	104.6689	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-7.8571	0.9995	61.7908	<.0001
mean_pain	1	-0.1076	0.0684	2.4728	0.1158
age	1	0.0413	0.0118	12.2649	0.0005
male	1	-0.0289	0.2302	0.0158	0.9000
e_score	1	0.4230	0.0523	65.3771	<.0001
oasis	1	0.0366	0.0150	5.9631	0.0146

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
mean_pain	0.898	0.785	1.027
age	1.042	1.018	1.066
male	0.971	0.619	1.525
e_score	1.527	1.378	1.691
oasis	1.037	1.007	1.068

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	78.6	Somers' D	0.572
Percent Discordant	21.4	Gamma	0.572
Percent Tied	0.0	Tau-a	0.060
Pairs	185640	c	0.786

Partition for the Hosmer and Lemeshow Test					
		X1_yr = 1		X1_yr = 0	
Group	Total	Observed	Expected	Observed	Expected
1	189	0	1.37	189	187.63
2	189	3	2.35	186	186.65
3	189	2	3.27	187	185.73
4	189	6	4.21	183	184.79
5	189	4	5.28	185	183.72
6	189	5	6.86	184	182.14
7	189	11	8.87	178	180.13
8	189	13	11.86	176	177.14
9	189	14	18.27	175	170.73
10	188	46	41.67	142	146.33

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
6.0183	8	0.6452

Sensitivity Model 4: Categorical Pain vs Hospital LOS

All CABG patients included

Number of Observations Read	1889
Number of Observations Used	1889

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	13724	2744.72524	93.79	<.0001
Error	1883	55103	29.26359		
Corrected Total	1888	68827			

Root MSE	5.40958	R-Square	0.1994
Dependent Mean	9.05966	Adj R-Sq	0.1973
Coeff Var	59.71064		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	5.70144	1.06297	5.36	<.0001
cat_pain	1	-1.70596	0.20636	-8.27	<.0001
male	1	-1.04945	0.29877	-3.51	0.0005
age	1	0.02149	0.01266	1.70	0.0900
e_score	1	1.14537	0.07384	15.51	<.0001
oasis	1	0.07046	0.01900	3.71	0.0002

Sensitivity Model 5: Categorical Pain vs 30-day Mortality

All CABG patients included

Number of Observations Read	1889
Number of Observations Used	1889

Response Profile		
Ordered Value	X30_day	Total Frequency
1	1	38
2	0	1851

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	374.103	327.406
SC	379.647	360.669
-2 Log L	372.103	315.406

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	56.6967	5	<.0001
Score	64.5808	5	<.0001
Wald	53.1142	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-6.9727	1.5104	21.3116	<.0001
cat_pain	1	-1.1138	0.2957	14.1909	0.0002
age	1	0.0213	0.0180	1.3982	0.2370
male	1	-0.0529	0.3617	0.0214	0.8836
e_score	1	0.3400	0.0788	18.6130	<.0001
oasis	1	0.0482	0.0243	3.9417	0.0471

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
cat_pain	0.328	0.184	0.586
age	1.022	0.986	1.058
male	0.948	0.467	1.927
e_score	1.405	1.204	1.640
oasis	1.049	1.001	1.100

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	80.4	Somers' D	0.607
Percent Discordant	19.6	Gamma	0.607
Percent Tied	0.0	Tau-a	0.024
Pairs	70338	c	0.804

Partition for the Hosmer and Lemeshow Test					
		X30_day = 1		X30_day = 0	
Group	Total	Observed	Expected	Observed	Expected
1	189	0	0.31	189	188.69
2	189	1	0.60	188	188.40
3	189	1	0.90	188	188.10
4	189	1	1.21	188	187.79
5	189	3	1.57	186	187.43
6	189	0	2.04	189	186.96
7	189	5	2.67	184	186.33
8	189	3	3.91	186	185.09
9	189	7	6.63	182	182.37
10	188	17	18.17	171	169.83

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
6.3800	8	0.6047

Sensitivity Model 6: Categorical pain vs 1-yr Mortality

All CABG patients included

Number of Observations Read	1889
Number of Observations Used	1889

Response Profile		
Ordered Value	X1_yr	Total Frequency
1	1	104
2	0	1785

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	807.244	701.692
SC	812.788	734.955
-2 Log L	805.244	689.692

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	115.5518	5	<.0001
Score	129.8715	5	<.0001
Wald	104.8178	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-7.8363	0.9896	62.7055	<.0001
cat_pain	1	-0.3017	0.1725	3.0581	0.0803
age	1	0.0411	0.0117	12.3074	0.0005
male	1	-0.0158	0.2305	0.0047	0.9454
e_score	1	0.4232	0.0522	65.6557	<.0001
oasis	1	0.0367	0.0150	5.9893	0.0144

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
cat_pain	0.740	0.527	1.037
age	1.042	1.018	1.066
male	0.984	0.626	1.547
e_score	1.527	1.378	1.691
oasis	1.037	1.007	1.068

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	78.7	Somers' D	0.574
Percent Discordant	21.3	Gamma	0.574
Percent Tied	0.0	Tau-a	0.060
Pairs	185640	c	0.787

Partition for the Hosmer and Lemeshow Test					
		X1_yr = 1		X1_yr = 0	
Group	Total	Observed	Expected	Observed	Expected
1	189	1	1.36	188	187.64
2	189	1	2.33	188	186.67
3	189	5	3.23	184	185.77
4	189	3	4.18	186	184.82
5	189	6	5.32	183	183.68
6	189	6	6.81	183	182.19
7	189	10	8.84	179	180.16
8	189	12	11.88	177	177.12
9	189	16	18.34	173	170.66
10	188	44	41.71	144	146.29

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
3.0321	8	0.9323

Sensitivity Model 7: Mean pain vs Hospital Length of Stay

Excluding in hospital mortality

Number of Observations Read	1867
Number of Observations Used	1867

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	12848	2569.60206	91.82	<.0001
Error	1861	52078	27.98408		
Corrected Total	1866	64926			

Root MSE	5.29000	R-Square	0.1979
Dependent Mean	9.01968	Adj R-Sq	0.1957
Coeff Var	58.64951		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	6.06474	1.05804	5.73	<.0001
mean_pain	1	-0.70050	0.07969	-8.79	<.0001
male	1	-0.95798	0.29392	-3.26	0.0011
age	1	0.01991	0.01256	1.58	0.1132
e_score	1	1.11301	0.07288	15.27	<.0001
oasis	1	0.06609	0.01871	3.53	0.0004

Sensitivity Model 8: Mean Pain vs 30-day Mortality

Excluding in hospital mortality

Number of Observations Read	1867
Number of Observations Used	1867

Response Profile		
Ordered Value	X30_day	Total Frequency
1	1	16
2	0	1851

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	186.166	172.675
SC	191.699	205.868
-2 Log L	184.166	160.675

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	23.4914	5	0.0003
Score	27.1882	5	<.0001
Wald	23.1706	5	0.0003

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-10.2879	2.4688	17.3652	<.0001
mean_pain	1	-0.2196	0.1777	1.5273	0.2165
age	1	0.0500	0.0295	2.8652	0.0905
male	1	0.0170	0.5518	0.0009	0.9755
e_score	1	0.3972	0.1162	11.6884	0.0006
oasis	1	0.0394	0.0358	1.2124	0.2709

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
mean_pain	0.803	0.567	1.137
age	1.051	0.992	1.114
male	1.017	0.345	2.999
e_score	1.488	1.185	1.868
oasis	1.040	0.970	1.116

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	79.7	Somers' D	0.593
Percent Discordant	20.3	Gamma	0.593
Percent Tied	0.0	Tau-a	0.010
Pairs	29616	c	0.797

Partition for the Hosmer and Lemeshow Test					
		X30_day = 1		X30_day = 0	
Group	Total	Observed	Expected	Observed	Expected
1	187	0	0.13	187	186.87
2	187	0	0.26	187	186.74
3	187	0	0.38	187	186.62
4	187	2	0.50	185	186.50
5	188	1	0.65	187	187.35
6	187	0	0.88	187	186.12
7	187	2	1.21	185	185.79
8	187	2	1.71	185	185.29
9	187	0	2.68	187	184.32
10	183	9	7.60	174	175.40

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
9.8622	8	0.2748

Sensitivity Model 9: Mean Pain vs 1-yr Mortality

Excluding in hospital mortality

Number of Observations Read	1867
Number of Observations Used	1867

Response Profile		
Ordered Value	X1_yr	Total Frequency
1	1	82
2	0	1785

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	674.905	591.678
SC	680.437	624.870
-2 Log L	672.905	579.678

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	93.2272	5	<.0001
Score	104.3845	5	<.0001
Wald	86.3856	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-9.1906	1.1351	65.5626	<.0001
mean_pain	1	0.0267	0.0738	0.1307	0.7177
age	1	0.0530	0.0135	15.5081	<.0001
male	1	-0.00300	0.2564	0.0001	0.9907
e_score	1	0.4467	0.0577	59.9324	<.0001
oasis	1	0.0309	0.0165	3.5163	0.0608

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
mean_pain	1.027	0.889	1.187
age	1.054	1.027	1.083
male	0.997	0.603	1.648
e_score	1.563	1.396	1.750
oasis	1.031	0.999	1.065

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	78.8	Somers' D	0.575
Percent Discordant	21.2	Gamma	0.575
Percent Tied	0.0	Tau-a	0.048
Pairs	146370	c	0.788

Partition for the Hosmer and Lemeshow Test					
		X1_yr = 1		X1_yr = 0	
Group	Total	Observed	Expected	Observed	Expected
1	187	1	0.99	186	186.01
2	187	1	1.75	186	185.25
3	187	2	2.46	185	184.54
4	187	3	3.26	184	183.74
5	187	6	4.20	181	182.80
6	187	1	5.35	186	181.65
7	187	11	7.02	176	179.98
8	187	13	9.26	174	177.74
9	187	8	14.23	179	172.77
10	184	36	33.49	148	150.51

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
11.9787	8	0.1522

Sensitivity Model 10: Categorical Pain vs Hospital LOS

Excluding in hospital mortality

Number of Observations Read	1867
Number of Observations Used	1867

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	12579	2515.76531	89.44	<.0001
Error	1861	52348	28.12873		
Corrected Total	1866	64926			

Root MSE	5.30365	R-Square	0.1937
Dependent Mean	9.01968	Adj R-Sq	0.1916
Coeff Var	58.80089		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	5.63343	1.05160	5.36	<.0001
cat_pain	1	-1.68014	0.20479	-8.20	<.0001
male	1	-0.92734	0.29477	-3.15	0.0017
age	1	0.02397	0.01253	1.91	0.0559
e_score	1	1.11908	0.07308	15.31	<.0001
oasis	1	0.06514	0.01875	3.47	0.0005

Sensitivity Model 11: Categorical Pain vs 30-day Mortality

Excluding in hospital mortality

Number of Observations Read	1867
Number of Observations Used	1867

Response Profile		
Ordered Value	X30_day	Total Frequency
1	1	16
2	0	1851

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	186.166	173.641
SC	191.699	206.834
-2 Log L	184.166	161.641

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	22.5254	5	0.0004
Score	26.1846	5	<.0001
Wald	22.6468	5	0.0004

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-10.7032	2.4542	19.0199	<.0001
cat_pain	1	-0.3439	0.4230	0.6607	0.4163
age	1	0.0531	0.0296	3.2079	0.0733
male	1	0.0324	0.5518	0.0035	0.9531
e_score	1	0.4126	0.1162	12.6022	0.0004
oasis	1	0.0375	0.0354	1.1222	0.2894

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
cat_pain	0.709	0.309	1.625
age	1.055	0.995	1.118
male	1.033	0.350	3.046
e_score	1.511	1.203	1.897
oasis	1.038	0.969	1.113

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	79.6	Somers' D	0.592
Percent Discordant	20.4	Gamma	0.592
Percent Tied	0.0	Tau-a	0.010
Pairs	29616	c	0.796

Partition for the Hosmer and Lemeshow Test					
		X30_day = 1		X30_day = 0	
Group	Total	Observed	Expected	Observed	Expected
1	187	0	0.15	187	186.85
2	188	0	0.28	188	187.72
3	188	1	0.40	187	187.60
4	187	1	0.53	186	186.47
5	187	0	0.69	187	186.31
6	187	1	0.92	186	186.08
7	187	2	1.23	185	185.77
8	187	1	1.72	186	185.28
9	187	2	2.66	185	184.34
10	182	8	7.42	174	174.58

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
3.4540	8	0.9027

Sensitivity Model 12: Categorical Pain vs 1-yr Mortality

Excluding in hospital mortality

Number of Observations Read	1867
Number of Observations Used	1867

Response Profile		
Ordered Value	X1_yr	Total Frequency
1	1	82
2	0	1785

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	674.905	591.769
SC	680.437	624.962
-2 Log L	672.905	579.769

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	93.1360	5	<.0001
Score	104.3571	5	<.0001
Wald	86.3552	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-9.1205	1.1300	65.1401	<.0001
cat_pain	1	0.0376	0.1909	0.0389	0.8436
age	1	0.0524	0.0134	15.2784	<.0001
male	1	-0.00484	0.2565	0.0004	0.9850
e_score	1	0.4444	0.0574	59.9633	<.0001
oasis	1	0.0312	0.0165	3.5868	0.0582

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
cat_pain	1.038	0.714	1.509
age	1.054	1.026	1.082
male	0.995	0.602	1.645
e_score	1.560	1.394	1.745
oasis	1.032	0.999	1.066

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	78.8	Somers' D	0.576
Percent Discordant	21.2	Gamma	0.576
Percent Tied	0.0	Tau-a	0.048
Pairs	146370	c	0.788

Partition for the Hosmer and Lemeshow Test					
		X1_yr = 1		X1_yr = 0	
Group	Total	Observed	Expected	Observed	Expected
1	187	1	1.00	186	186.00
2	187	1	1.75	186	185.25
3	187	2	2.47	185	184.53
4	187	3	3.26	184	183.74
5	187	6	4.20	181	182.80
6	187	1	5.36	186	181.64
7	187	9	7.01	178	179.99
8	187	15	9.27	172	177.73
9	187	9	14.22	178	172.78
10	184	35	33.48	149	150.52

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
11.3564	8	0.1823

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*

Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any pre-specified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	3-4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	4
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4-5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*4-5*
Bias	9	Describe any efforts to address potential sources of bias	9
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6
		(*b*) Describe any methods used to examine subgroups and interactions	6
		(*c*) Explain how missing data were addressed	6
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	NA

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
		(*e*) Describe any sensitivity analyses	6
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	NA
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	6-7, Table 1
		(b) Indicate number of participants with missing data for each variable of interest	NA
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	8, Table 2
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Table 2
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	7
Discussion			
Key results	18	Summarise key results with reference to study objectives	7
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	9
Generalisability	21	Discuss the generalisability (external validity) of the study results	9
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	13

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.
Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.