

Development and validation of an electronic postoperative morbidity score (EPOMS)

Daniel J Stubbs BMBCh [1], Jessica L Bowen MBChB [1] Rachel C Murray MBChB [1]
Fay J Gilder MB BS [1] Roman Romero-Ortuno PhD [2] Richard Biram MBBS [2],
David K Menon FMedSci [1] Ari Ercole PhD [1]

1: University Division of Anaesthesia, Addenbrooke’s Hospital, Hills Road, Cambridge,
CB2 0QQ

2: Department of Medicine for the Elderly, Addenbrooke’s Hospital, Hills Road,
Cambridge, CB2 0QQ

Corresponding author: Dr DJ Stubbs: University of Cambridge, University Division of
Anaesthesia, Addenbrooke’s Hospital, Hills Road, Cambridge, UK, CB2 0QQ, Email:
djs225@cam.ac.uk, Telephone: 01223 274446

Funding Disclosures: None

Conflicts of Interest: None

Author contributions

Stubbs DJ: This author was involved in concept, data extraction, manuscript preparation
and led statistical analysis.

Bowen JL: This author helped with data extraction, interpretation and manuscript review

Murray R: This author helped with data extraction, interpretation and manuscript review

Gilder F: This author helped with concept, patient selection, and interpretation of findings

Biram R: This author helped with concept, patient selection, and interpretation of findings

Romero-Ortuno R: This author helped with concept, statistical review, and interpretation.

Menon DK: This author helped with interpretation and manuscript review

Ercole A: This author helped with, concept, interpretation, manuscript review, design of statistical analysis and assessment.

Abstract: 255 words

Introduction: 394 words

Discussion: 1598 words.

Total (excl references and abstract): 3702 words

Abstract:

Background: Electronic health records (EHR) are being adopted due to numerous potential benefits. This requires the development of objective metrics to characterize morbidity, comparable to studies performed in centers without an EHR. We outline development of an electronic version (E-POMS) of the postoperative morbidity score (POMS) for integration into our EHR.

Methods: 203 frail patients who underwent elective surgery were reviewed. We retrospectively defined POMS morbidity on postoperative day 3 (D3). We also recorded potential electronic surrogates for morbidities that could not be easily extracted in an objective format. We compared discriminative capability (area under the receiver operator curve - AUC) for patients having prolonged length of stay (pLOS) or complex discharge requirements (CD).

Results: 139 patients (68%) had morbidity in at least one POMS domain. Initial electronic surrogates were overly sensitive, identifying 173 patients (84%) as having morbidity. We refined our definitions using backwards logistic regression against 'gold-standard' POMS. The final E-POMS differed from the initial version in its definition of cardiac and neurological morbidity. There was no significant difference in discriminative capability between E-POMS and POMS for either outcome (AUC 0.66 v 0.66 for CD, AUC 0.66 v 0.67 for pLOS, $p > 0.05$ for both). Patients with POMS or E-POMS defined morbidity on D3 had increased risk of prolonged length of stay ($p < 0.001$ for both)

Conclusions: We present a variant of POMS based on objective electronic metrics. Discriminative performance appeared comparable to gold-standard definitions for discharge outcomes. E-POMS may allow characterization of morbidity within our EHR but further work is required to assess external validity.

Key Points:

Question: Can surrogates for a validated post-operative morbidity score be extracted from an electronic patient record?

Findings: Use of electronic surrogates for previously published definitions had comparable discriminative power for the identification of patients at risk of clinically relevant discharge outcomes.

Meaning: As increasing numbers of hospitals adopt electronic health records, local development of such measures is vital to assess performance and characterize patient outcome.

Introduction

Defining postoperative morbidity is vital for measuring patient outcomes as well as quantifying the risks of specific interventions. Risk scores are often used to quantify such information for patients and clinicians¹. Importantly, how complications are defined may lead to differences in reported incidences;² as such, the use of standardized outcome measures is to be advocated. Ongoing work is looking to develop a consensus set of outcomes for perioperative medicine³.

The 'Perioperative Quality Improvement Program' (PQIP) in the United Kingdom includes identification of operative morbidity as a core indicator of patient outcome⁴.

PQIP defines postoperative morbidity using the "Postoperative Morbidity Score" (POMS). This records morbidity across nine organ system domains (**Table 1**). Initially studied in a surgically heterogeneous population of 438 individuals in the United States⁵, the score was calculated prospectively on days 3, 5, 8, and 15. Calculation was based on medical notes, patient assessment, and discussion with caregivers. A modification of the score was assessed in the United Kingdom in 2007 in a mixed surgical cohort⁶.

Subsequent work demonstrated that retrospective collection of POMS data was non-inferior⁷.

In perioperative medicine, large datasets are required to draw inferences due to heterogeneity of practice. This is only feasible if data extraction can be automated. Many hospitals are acquiring integrated electronic health record systems (EHR). Such systems

offer improved legibility, accuracy, and opportunities for perioperative research⁸. The epidemiological advantages extend beyond the direct realms of anesthetic practice; information has been used to characterize trends in management of coronary artery disease in patients presenting for surgery⁹. These benefits rely on information being coded in a readily extractable, objective format. It is vital to identify markers within the EHR which can replace laborious hand-searching of free text components. Unfortunately, the POMS score contains elements which require note review for interpretation. An electronic version is therefore needed to facilitate ‘big data’ analysis. Any derived score needs validation against the original version to ensure interoperability.

We aimed to generate an ‘electronic variant of the POMS’ (“E-POMS”) that could be integrated into our EHR for our ongoing quality improvement work. Using a dataset of frail patients undergoing elective surgery we undertook a process of notes review to evaluate morbidity as defined by POMS. Alongside we extracted electronic markers of morbidity that we would explore as surrogates. We compared E-POMS to the original POMS for identifying discharge outcomes of interest.

Methods

Approval

Work was conducted as part of a retrospective service evaluation and quality improvement project. Approval was granted by our hospital trust’s Patient Safety

Department (local reference PRN6715). Being as this was a retrospective analysis the patient safety department made no request for the gaining of informed consent.

Setting:

Addenbrooke's Hospital is a tertiary surgical center located in Cambridge, United Kingdom. A hospital wide EHR with an integrated Anesthesia Information Management System (AIMS) (Epic Systems Inc, Verona-WI) has been in use since 2014. This provides readily available information on intraoperative events as well as pre-operative and postoperative information. Laboratory results, observations, comorbidities, and prescriptions are all exportable, and linked to specific hospital encounters. Prior to elective surgery, patients are routinely reviewed in a multi-disciplinary pre-assessment clinic if they are identified as 'frail' (Rockwood clinical frailty scale - CFS of 4 or more)¹⁰. Here they are reviewed by a consultant anesthetist and consultant geriatrician alongside a physiotherapist and occupational therapist.

Patient population

We included all patients who had attended our multi-disciplinary pre-assessment clinic from January 2016 to the end of June 2017. Those that did not proceed to surgery, had yet to undergo their operation or were less than 30 days from surgery were excluded. We identified morbidity on postoperative day 3 (D3). This decision was based on previous work demonstrating highest incidence of morbidity at this time-point,¹¹ therefore we only included patients with a length of stay of 3 days or longer.

Definitions:

We retrospectively searched our EHR on D3 for evidence of complications as defined by the POMS. Certain modifications (noted in **Table 1** and referenced) were made. This involved the searching of electronic observation charts, medical notes, and drug charts. POMS domains were scored as present or absent. We prospectively defined potential electronic surrogates for POMS domains (**Table 1**). In many cases these were identical to the literature standard. Surrogates were chosen with the intention of mapping to the original definition (e.g. furosemide prescription as a surrogate for pulmonary edema). The presence of these electronic definitions were also extracted. With the following caveats, all data were extracted on calendar D3; postoperative creatinine samples were included from D2 to D4, further operations for wound exploration were included in a 12-hour period on either side of postoperative day 3. Literature standard cardiovascular complications were included in the 24-hour period running from D3 retrospectively into D2. Data for the electronic definitions was recorded ‘blind’ without interpretation in light of other information (i.e. if no urine output was recorded as the patient was not catheterized, then this was recorded as a urine output of < 500 ml/24hrs). When no suitable postoperative creatinine was available the pre- to postoperative creatinine ratio was recorded as 0 and included in analysis. A positive troponin test was defined as a result above the normal range for our institution (Troponin I 0-5ng/L).

Outcomes

We assessed performance of both POMS and E-POMS against predefined outcomes.

Firstly ‘complex discharge’ (CD: a composite of whether an individual required discharge to another institution or their own home with an altered care package) or ‘prolonged length of stay’ (pLOS; a hospital stay of 8 days or more).

Statistical analysis

All analysis was performed in R (R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <http://www.r-project.org/>) Comparison of initial distributions of POMS and E-POMS was conducted using the Wilcoxon signed rank test. Scoring systems were compared as; a binary outcome (any complication yes/no), an additive score (0-9, depending on the number of domains an individual was positive for) and with regression of all domain predictors against each outcome. For refinement of electronic definitions, candidate surrogates were regressed against a ‘gold standard’ outcome as defined by the POMS criteria. Backwards logistic regression was performed using the ‘step’ function in R. Simplified models were compared using analysis of deviance (likelihood ratio) testing to ensure that simplification had caused a non-significant change in explained variance. Impact of this process was assessed by the generation of classifier metrics including sensitivity, specificity, and positive and negative predictive values for the ability of initial, and refined EPOMS scores to identify differing levels of POMS morbidity. Model discrimination was assessed using area under the receiver operator curve (AUC). The final score was compared to POMS for the discrimination of our outcomes of interest.

Calibration was assessed using cross-calibration plots. Here morbidity predictions generated by logistic regression models based on literature POMS and E-POMS definitions were correlated against each other. The equation of the resulting correlation line allows for an interpretation of systemic offset between the two models (the intercept, 'calibration in the large') as well as the calibration slope. An ideal model would have values of 0 and 1 for each of these respectively ¹¹. Finally, Kaplan-Meier curves were constructed for length of stay in days against those patients found to be POMS or EPOMS positive/negative. Significance of survival analysis was performed using the log-rank test. In all circumstances, unless otherwise stated, a P value of 0.05 was taken as the threshold for significance.

Results

207 patients from our initial cohort of 332 had a length of stay of 3 days or more. Of these, 4 died as an inpatient and were excluded from further analysis leaving a final dataset of 203 patients. Median length of stay (LOS) was 7 days (interquartile range 5-11). 48 (24%) had a complex discharge. 94 patients (46%) had pLOS. The majority (112 - 55%) underwent orthopedic surgery. 24 (12%) underwent general surgical procedures, 20 (10%) urological procedures and 19 (9%), vascular surgery. Mean (standard deviation) age was 80 (7.8). 107 (53%) were female.

Incidence of morbidity

On D3 139 (68%) patients had morbidity in at least one POMS domain. Initial electronic definitions identified 173 (84%) patients with morbidity in at least one domain.

Distribution of incidence of both scores is demonstrated in **Figure 1**. The difference in occurrence of morbidity was statistically significant ($p < 0.001$). **Supplementary Table 1** identifies mismatch between our initial electronic score and POMS.

Domains where our new score is based entirely on the POMS score demonstrate exact matching. However our surrogates for cardiac, renal, and neurological domains appeared to be overly sensitive, identifying more patients than POMS. Using anti-emetic prescriptions alone we missed three patients with gastrointestinal morbidity. Being as this modification had previously been made to the POMS score⁶ we did not consider this domain for refinement.

Performance of initial EPOMS

We assessed the performance of our initial score for identifying patients with any level of POMS defined morbidity (i.e. any domain positive). Test performance metrics are shown in **Table 2**. Our initial score demonstrated high sensitivity (99.3%) but was relatively non-specific (45.3%).

Discriminative performance of initial score

Presence of any POMS morbidity on D3 had an AUC of 0.61 [95% Confidence Interval 0.54-0.67] for discriminating between patients with pLOS. For initial electronic definitions AUC was 0.53 [0.48-0.58] ($p < 0.01$). For CD the AUCs were 0.57 [0.50-

0.64] and 0.53 [0.48- 0.58] for POMS, and electronic definitions, respectively, indicating non-significant discriminative capability ($p = 0.29$). We explored discrimination of definitions when employed as an ordinal score (0-9) or as a regression model of all domains. Highest AUCs were found with a regression model of all domains: 0.67[0.57-0.76] (POMS) v 0.68[0.59-0.76] (electronic) for CD and 0.67[0.60-0.75] (POMS) v 0.69[0.62-0.76] (electronic) for pLOS.

Refining domain definitions

We sought to refine the initial definitions of cardiac, neurological, and renal morbidity in an effort to improve overall E-POMS performance. This was done using backwards regression of our initial surrogates in a specific domain against positivity in the POMS domain of interest. The end results of this process are the domain definitions presented in Table 3 which form our final E-POMS score. New domain definitions had AUCs of 0.68[0.56-0.81] (cardiac), 0.86 [0.75-0.97] (neurological), and 0.97 [0.92-1.00] (renal) for the discrimination of individuals classed as positive in the corresponding POMS domain. In all cases, analysis of deviance testing of simplified models indicated a non-significant change in explained variance ($P > 0.05$ for all). Classifier performance of the final score is shown in Table 2. Changes in positivity across domains compared to POMS and our initial score is shown in Supplementary Table S1.

Final E-POMS

Correlation of final E-POMS and POMS totals was performed, this indicated good correlation (Spearman's $\rho = 0.88$; $p < 0.0001$). We formed logistic regression models using all domains of final E-POMS. This was compared against a logistic regression model consisting of all domains of POMS. Receiver operator curves (ROC) for both outcomes are demonstrated in **Figure 2**. There was no significant difference in discriminative capability between final E-POMS and POMS for either outcome (AUC 0.66 [0.57-0.75] v 0.66[0.57- 0.76] for complex discharge, AUC 0.66 [0.59-0.74] v 0.67 [0.60-0.75] for pLOS, $p > 0.05$ for both). Cross-calibration plots of generated predictions from both models for both outcomes were generated. These demonstrated little offset (intercepts 0.042 and 0.036 for pLOS and CD respectively). Calibration slopes for both were < 1 (0.91 and 0.85) indicating a degree of optimism in predictions. As a binary classifier the final variant of EPOMS demonstrated improved specificity (78.1%) and positive predictive value (90.1%) for identification of POMS defined morbidity when compared to our initial score (**Table 2**). This trend was seen for the identification of all levels of morbidity (with a corresponding increase in overall accuracy) but, at higher levels was associated with a decreasing level of sensitivity. When assessed as an additive score (i.e. total domains positive on a scale of 0-9), agreement between the total POMS and total EPOMS score improved from 45.8% using initial definitions to 69.5% using our final version.

We assessed the impact of accumulating morbidity on discrimination of CD and pLOS by identifying patients with total POMS and EPOMS scores of 2 or more, or 3 or more. All

resulting AUCs were <0.57 with no significant difference between POMS and EPOMS ($P > 0.05$).

Kaplan-Meier curves for length of stay dichotomized by D3 morbidity defined by either POMS or E-POMS scores of ≥ 1 versus no morbidity are shown in **Figure 3**. Both identified patients with an increased hazard of longer length of stay (from D3 onwards) ($P < 0.01$ log-rank hazard score) Even after removal of an outlying patient with a LOS > 100 days these findings remained significant (data not shown).

Discussion

The POMS is a multi-dimensional outcome score that has been employed in different healthcare settings⁵⁻⁶ and is currently being used as an outcome metric in a national quality improvement program (PQIP) being run by the Royal College of Anaesthetists in the United Kingdom⁴. We present a version of the POMS whose constituents can be readily extracted from our EHR and whose performance appeared comparable to the original for discrimination of clinically relevant discharge outcomes. Our final list of variables is only significantly different from the original POMS within the cardiac and neurological domains. External validity cannot be inferred from a single center study but the approach we have taken may be of interest to those looking to develop local EHR metrics of morbidity.

There are certain caveats that need to be discussed. Firstly, the score has been developed on a modestly sized dataset derived from a single center using retrospective data. Since

calculation of the original POMS requires note review, this limits the size of sample that can be realistically assessed. We utilized a single time point to make the amount of manual note-searching tractable. Given its similarity to the original definitions of POMS, it is likely the score will have similar discriminative capability when applied at different time points, however this awaits confirmation. Importantly, this introduces an element of bias into our results as it is possible we have selected complications occurring due to distinct causes than at earlier or later timepoints. For instance presence of oxygen therapy on post-operative day zero or one may simply reflect residual anaesthetic effect. Use of the score at distinct temporal points and with reference to accumulated duration of morbidity is necessary to explore the impact and importance of any bias resulting from our methods. Secondly, although at a population level the discriminative capability of the two scores is not significantly different, this will not necessarily hold true at the level of an individual. This will mean there will be individuals classed as morbidity 'positive' or 'negative' differently by these, and other surrogate markers. To some extent this could be mitigated by co-linearity inherent within POMS; for instance a patient suffering from pulmonary edema is likely to be receiving oxygen and thus will score for both the cardiac and respiratory domains. These differences may be less relevant in analyses that involve patient populations, but the clinical expectations of outcome that derive from a given POMS score in an individual may not translate faithfully when an EPOMS score is used at its current level of development. Consequently, we would urge caution in this context

until further experience and refinement of EPOMS methodology allow more reliable clinical inferences.

Assessment of the classifier performance of EPOMS for the identification of POMS defined morbidity demonstrates important information on our methodology, the utility of the score as a screening tool, and the need for validation in larger data sets. Broadly the process of domain refinement improved the specificity, and thus overall accuracy, of our final version of EPOMS (**Table 2**). This demonstrates that our methodological process of definition refinement was valid and is likely due to a reduction in the number of ‘false positives’ driven by the overly sensitive cardiovascular domain. The final scores’ increased specificity and positive predictive value compared to the initial variant highlight its improved (but not perfect) ability to identify the individual with morbidity, as defined by POMS. When used to identify patients with multiple morbidities there is a tail off in sensitivity in the revised version – this could be due to the refinement of the cardiovascular domain, and lower numbers of patients suffering ‘rarer’ complications for which our definitions require further refinement. This can only be tested in a larger dataset. As a binary classifier the final version of EPOMS demonstrates excellent sensitivity (>99%) and reasonable specificity (78%). The high sensitivity means that it would be well suited for the screening of patients at risk of deterioration, as very few (less than one in a hundred) patients with POMS defined morbidity would be missed.

Finally, our results indicate that certain domains (particularly cardiac) require refinement. Performance in the final model may be improved due to co-linearity but it is possible that

the introduction of other surrogates further removed from the original POMS (e.g. new prescriptions of cardiac medications) may offer even better performance. Whereas our cardiac definitions are likely to be overly sensitive (arguably of benefit if a score were to be used to identify deteriorating patients to prevent ‘failure to rescue’) the criteria for wound infection are likely insensitive. Our chosen definition (need for an operation to explore the wound) is in keeping with the original POMS definition but will only identify a minority of very severe cases. However, we hypothesise that in a larger dataset patients with lesser degrees of surgical site infection would be identified by the fever and antibiotic use encapsulated within the ‘Infectious’ domain definitions.

We explicitly aimed to identify surrogate markers that could be extracted in an unambiguous format, and which are likely to be EHR *platform agnostic*. We also sought markers that did not require other information for their interpretation. For instance, the use of furosemide prescriptions as a surrogate for pulmonary edema identified a number of patients receiving it chronically, thus its blanket inclusion would have worsened our specificity. Definitions were chosen so our score could, in theory, be integrated within an EHR. Having EPOMS embedded within an EHR raises the possibility of using it to track patient outcome in ‘real-time’ to identify patients requiring medical input. This could be a method for preventing ‘failure to rescue’.¹² However, due to difficulty with discriminating down to the level of individual organ systems, the exact format and thresholds to prompt such review can only be established from prospective work.

Others have already looked at identifying patients with a known disease state using information extractable from a primary care electronic record. They used markers (including prescriptions, blood results, body mass index) to develop algorithms that could identify cases of diabetes that had not been accurately coded within the EHR¹³. This work is analogous to our approach where a disease state (occurrence of a specific organ dysfunction) has been recorded in free text but not in a manner that can be readily identified on searching of the EHR.

The markers that we have extracted to generate E-POMS are based on practice in our institution. Other hospitals may have different or better surrogates that could be used. One marker worth considering is our use of a recorded ‘specialling scoring tool’ (SST) as a surrogate for the confused patient. Our hospital uses the SST to aid in the identification of patients requiring higher intensity nursing for behavioural, cognitive, environmental, or psychological reasons. We used the presence of a completed score (of any value) as a marker of needing closer nursing for delirium. It is possible that use of specific components of the score, rather than its completion *per se*, may offer greater specificity.

Within the United Kingdom, the National Institute for Health and Care Excellence (NICE) mandate that all hospitals have a mechanism for recording the need for higher intensity nursing¹⁴. Other institutions may record this in a different format but we hypothesize it will likely be recorded in a comprehensive EHR. Other hospitals may already record validated delirium screening tools, we have, (subsequent to this analysis), implemented a delirium screening tool (‘The 4AT’)¹⁵ for patients aged over 65. This

would theoretically offer a greater degree of both specificity and sensitivity by identifying those with hypoactive delirium in a way that our use of the SST may not. Its inclusion would be an easy modification to our tool and, in fact, would simply reflect an electronic record of the original POMS criteria rather than being a true surrogate.

It is important to note that retrospective application of the POMS itself may miss patients with significant morbidity. For instance, identification of patients with neurological complications does not require a formal delirium screening tool, merely documentation of ‘confusion, coma or delirium’. Thus patients with hypoactive or less severe hyperactive delirium may be missed retrospectively. As such, certain definitions within the POMS are likely to be specific but not sensitive.

Mathematically, it would have been possible to simplify both EPOMS and POMS during the regression process. We have intentionally avoided this for two reasons. Firstly, we aimed to develop a score that was recognizable to clinicians who may have already encountered the POMS. Consequently, we elected to keep domain number and classifications the same. Secondly, we recognized that the incidence of specific complications (e.g. focal neurology) within our sample was low. Therefore, although the exclusion of “CT Head” had minimal mathematical impact, this might not be the case in a larger sample. This lack of simplification raises the risk of model overfitting due to our event per variable ratio (EPV)¹⁶. Overfitting describes a model which loses its generalizability by being too closely tailored to the data it has been developed from. An EPV of ten is often advocated to minimize this risk¹⁶. For the prediction of pLOS our

EPV is ten (nine predictors against 94 outcomes) but for CD is closer to five (9 predictors against 48 outcomes). That our calibration slopes for both outcomes had gradients of <1 can also be interpreted as a sign of overfitting¹¹.

Conclusion

We present a variant of POMS whose constituents are readily extractable from an EHR. Surrogate markers were based on original definitions and the resulting score had comparable discriminative power to POMS for the prediction of discharge related outcomes when calculated on D3.

Acknowledgements

With many thanks to Ms Sarah Lester (Occupational Therapist, PRIME clinic, Addenbrooke's Hospital, Cambridge) for maintenance of clinic attendance database.

References

1. Prytherch, Whiteley, Higgins, Weaver, Prout, and Powell. POSSUM and Portsmouth POSSUM for Predicting Mortality. *Br J Surg*. 1998; 85: 1217–20. doi:10.1046/j.1365-2168.1998.00840.x.
2. Taylor JS, Marten CA, Potts KA et al. What Is the Real Rate of Surgical Site Infection?. *J Oncol Pract*. 2016; 12: e878–e883. doi:10.1200/jop.2016.011759.
3. Myles, PS., Grocott MPW, Boney O, and S.R. Moonesinghe. Standardizing End Points in Perioperative Trials: towards a Core and Extended Outcome Set. *Br J Anaesth* 2016; 116: 586–89. doi:10.1093/bja/aew066.
4. The Royal College of Anaesthetists. Perioperative Quality Improvement Programme: Introduction Pack. [https://pqip.org.uk/FilesUploaded/PQIP Introduction Pack v0.1 10.03.17.pdf](https://pqip.org.uk/FilesUploaded/PQIP%20Introduction%20Pack%20v0.1%2010.03.17.pdf). Accessed July 16th 2018
5. Bennett-Guerrero, E., Welsby I, Dunn T.J. et al. The Use of a Postoperative Morbidity Survey to Evaluate Patients with Prolonged Hospitalization After Routine Moderate-Risk, Elective Surgery. *Anesth Analg*. 1999; 89: 514–19. doi:10.1213/00000539-199908000-00050.
6. Grocott, M.P.W., Browne J.P., Van der Meulen J et al. The Postoperative Morbidity Survey Was Validated and Used to Describe Morbidity after Major Surgery. *J Clin Epidemiol* 2007; 60: 919–28. doi:10.1016/j.jclinepi.2006.12.003.

7. Patel, Amour B.U., Anna Reyes, and Gareth L. Ackland. Non-Inferiority of Retrospective Data Collection for Assessing Perioperative Morbidity”. *PeerJ* 2015; 3: e1466. doi:10.7717/peerj.1466.
8. Stabile M, and Lebron C. Review Article: The Evolving Role of Information Technology in Perioperative Patient Safety”. *Can J Anaesth* 2012; 60: 119–26. doi:10.1007/s12630-012-9851-0.
9. Vigoda MM, Rodriguez LI, Wu E et al. The Use of an Anesthesia Information System to Identify and Trend Gender Disparities in Outpatient Medical Management of Patients with Coronary Artery Disease. *Anesth Analg*. 2008; 107: 185–92. doi:10.1213/01.ane.0000289651.65047.3b.
10. Canadian Study for Health and Ageing (2008 revision): Clinical Frailty Scale Retrieved from:<http://camapcanada.ca/Frailtyscale.pdf>. Accessed July 16th 2018.
11. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the Performance of Prediction Models: A framework for traditional and novel measures. *Epidemiology*. 2010; 21: 128–38. doi:10.1097/ede.0b013e3181c30fb2.
12. Ghaferi, AA, Birkmeyer JD, and Dimick JB. Hospital Volume and Failure to Rescue With High-Risk Surgery. *Med Care*. 49: 1076–81. doi:10.1097/mlr.0b013e3182329b97.

13. Bagheri A, Sadek A, Chan T, Khunti K, and de Lusignan S. Using Surrogate Markers in Primary Electronic Patient Record Systems to Confirm or Refute the Diagnosis of Diabetes. *Inform Prim Care*. 2009; 17: 121–29.

14. National Institute for Health & Care Excellence (NICE). Safe staffing for nursing in adult inpatient wards in acute hospitals [SG1] available at: <https://www.nice.org.uk/guidance/SG1/chapter/1-Recommendations> accessed on 8th July 2018

15. O’Sullivan D, Brady N, Manning E et al. Validation of the 6-item cognitive impairment test and the 4AT test for combined delirium and dementia screening in older Emergency Department attendees. *Age Ageing*, 2018; 47: 61-68. doi: 10.1093/ageing/afx149.

16. Moons KG, Joris AH. de Groot WB et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLoS Med* 2014; 11: e1001744. doi:10.1371/journal.pmed.1001744.

Domain	POMS	Electronic Surrogates
Pulmonary	Requirement for supplemental oxygen or other respiratory support	As per POMS
Infectious	On antibiotics or temperature of 38°C or higher in last 24 hrs	As per POMS
Renal	Presence of oliguria (<500ml/24hrs), rise in serum creatinine of > 30% from baseline	As per POMS
Haematological	Requirement for blood, platelets, FFP, or cryoprecipitate in last 24 hrs	As per POMS
Pain	Wound pain requiring parenteral opiates or regional anesthesia	Active patient controlled analgesic or regional anaesthetic infusion (epidural/wound catheter) prescription
Gastrointestinal*	Unable to tolerate an enteral diet due to nausea, vomiting and	Anti-emetic administration

	abdominal distension OR use of an anti-emetic ⁶	
Cardiovascular*	Test or therapy in last 24 hrs for	HR > 100 bpm, SBP < 100 mmHg, furosemide prescription, positive troponin test
	myocardial infarction or ischaemia, hypotension requiring drug therapy or fluid > 200ml/hr, atrial or ventricular arrhythmia or pulmonary oedema	
Neurological*	New focal deficit, coma/confusion/delirium	Request for CT Head scan OR recorded specialling screening tool
Wound*	Wound dehiscence requiring surgical exploration or drainage of pus from the wound	Further operation performed

Table 1: Definitions of morbidity according to the Postoperative Morbidity Score (POMS) and our initial electronic surrogates. Asterisks represent domains where these definitions differ. BPM = beats per minutes. FFP = Fresh Frozen Plasma

	Versus POMS of 1 or more	
	Initial EPOMS	Final EPOMS
Sensitivity (%)	99.3	99.3
Specificity (%)	45.3	78.1
NPV	96.7	98.0
PPV	79.8	90.1
Accuracy (%)	82.3	92.6
	Versus POMS of 2 or more	
	Initial EPOMS	Final EPOMS
Sensitivity (%)	95.9	90.4
Specificity (%)	63.9	85.4
NPV	96.5	94.1
PPV	59.8	77.7
Accuracy (%)	75.4	87.2
	Versus POMS of 3 or more	
	Initial EPOMS	Final EPOMS
Sensitivity (%)	91.7	83.3
Specificity (%)	84.4	94.0
NPV	97.9	96.3
PPV	55.9	75.0

Accuracy (%)	85.7	92.1
--------------	------	------

Table 2: Test performance metrics for EPOMS at identifying presence of any POMS morbidity before (Initial EPOMS) and after (Final EPOMS) refinement of domain definitions. Classifier performance given for the identification of: any complication, 2 or more complications, 3 or more complications. Full confusion matrices used in calculations can be found in Supplementary Table S2 (A-F). NPV = Negative Predictive Value, PPV = Positive Predictive Value.

Domain	Final E-POMS definition
Pulmonary	Supplemental oxygen recorded OR respiratory support recorded
Infectious	On antibiotics or temperature of 38°C or higher in last 24 hrs
Renal*	Rise in serum creatinine of > 30% from baseline
Haematological	Requirement for blood, platelets, FFP or cryoprecipitate in last 24 hrs
Pain	Active patient controlled analgesic or regional anaesthetic infusion prescription
Gastrointestinal	Anti-emetic administration
Cardiovascular*	HR > 100bpm OR positive troponin

Neurological*	Recorded specialising scoring tool OR CT
	Head
Wound	Further operation

Table 3 Final definitions for our derived electronic POMS (E-POMS), FFP: Fresh Frozen Plasma. [*Indicates domain definition modified from original POMS definitions.](#)

Figure Legends

Figure 1: Histograms demonstrating distribution of morbidity in our patient population using initial definitions of an electronic score (A) or using previously published definitions for POMS (B).

Figure 2: Receiver Operator Curves (ROC) demonstrating discriminative capability of POMS and EPOMS for the identification of complex discharge (A) and prolonged LOS (B). For AUC values see text

Figure 3: Kaplan-Meier curves demonstrating length of stay for patients dichotomised by the presence of absence of comorbidity on D3 using: A, EPOMS definitions and B, POMS definitions. Tables demonstrate number of patients each category at each time point.

Supplementary Material Legends

Supplemental Table 1: Difference in incidence of morbidity using POMS

(Postoperative Morbidity Score) or our initial electronic markers

Supplementary Tables 2 A-F: Confusion matrices assessing performance of initial electronic definitions (tables **A, C, E**) or final EPOMS definitions (tables **B, D, F**) for the identification of differing degrees of POMS defined morbidity (total POMS score: ≥ 1 , ≥ 2 , ≥ 3). **A-B:** Performance of both definitions when 1 or more domains is positive. **C-D:** when 2 or more domains positive. **E-F:** when 3 or more domains positive. TP = True Positive, FP = False Positive, FN= False Negative, TN = True Negative

Calculations used:

Sensitivity = $TP/(TP+FN)$, Specificity = $(TN/(TN+FP))$, Negative Predictive Value = $(TN/(TN+FN))$, Positive Predictive Value = $(TP/(TP+FP))$

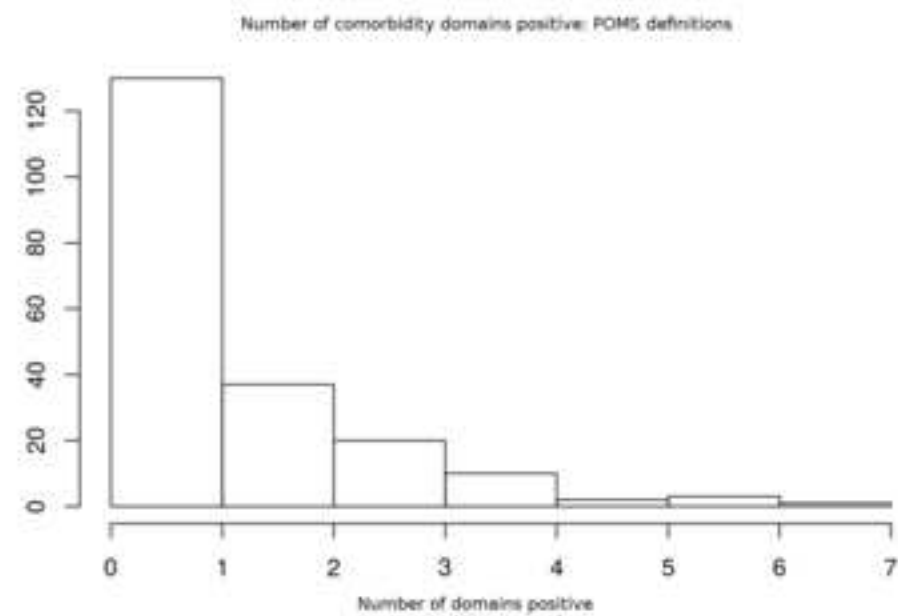
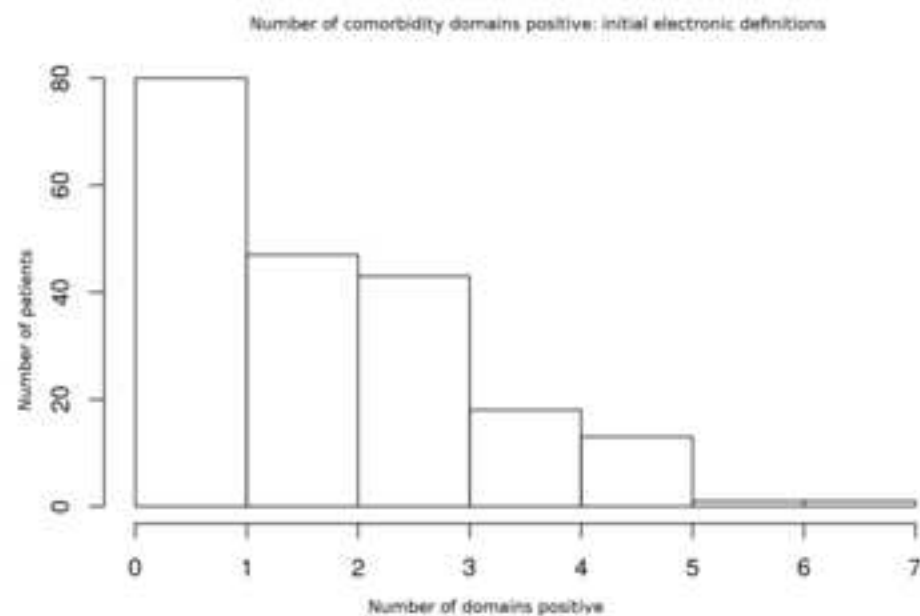
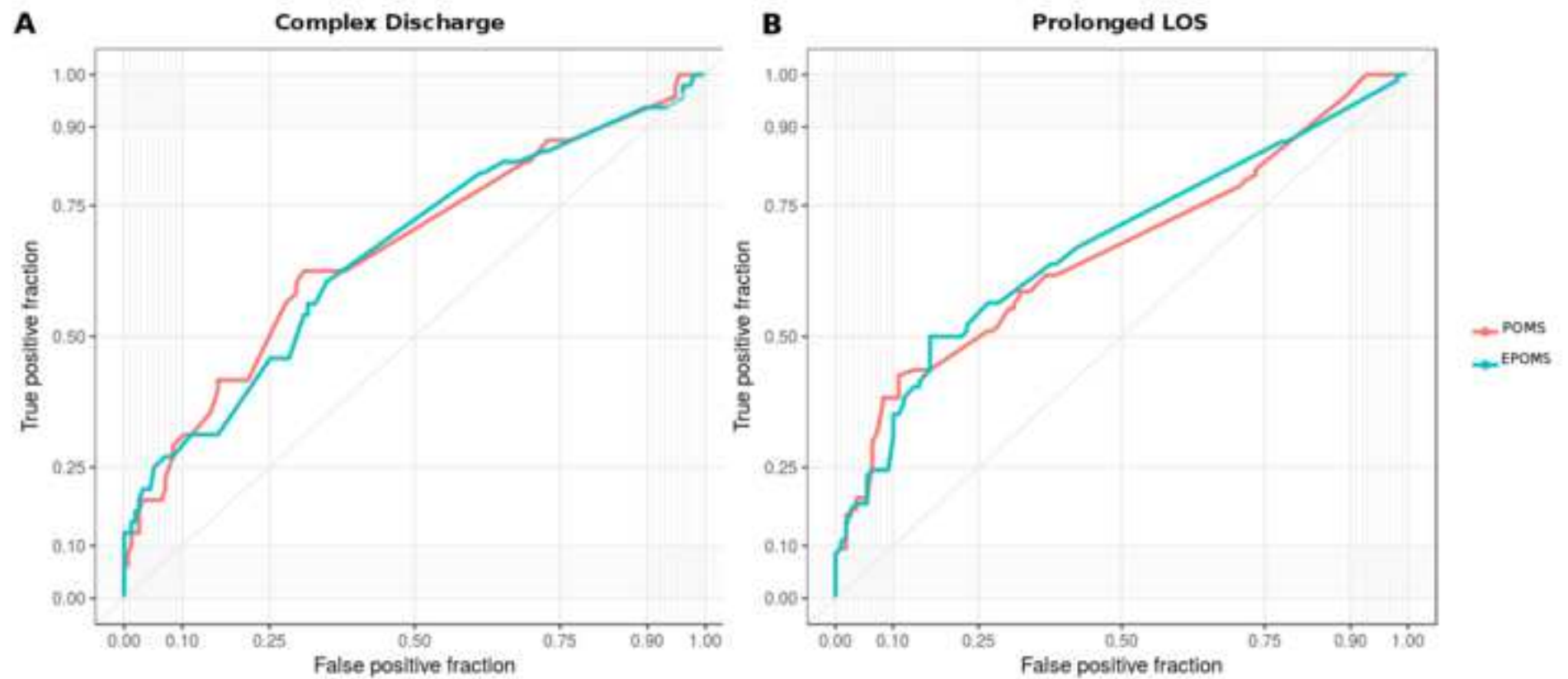
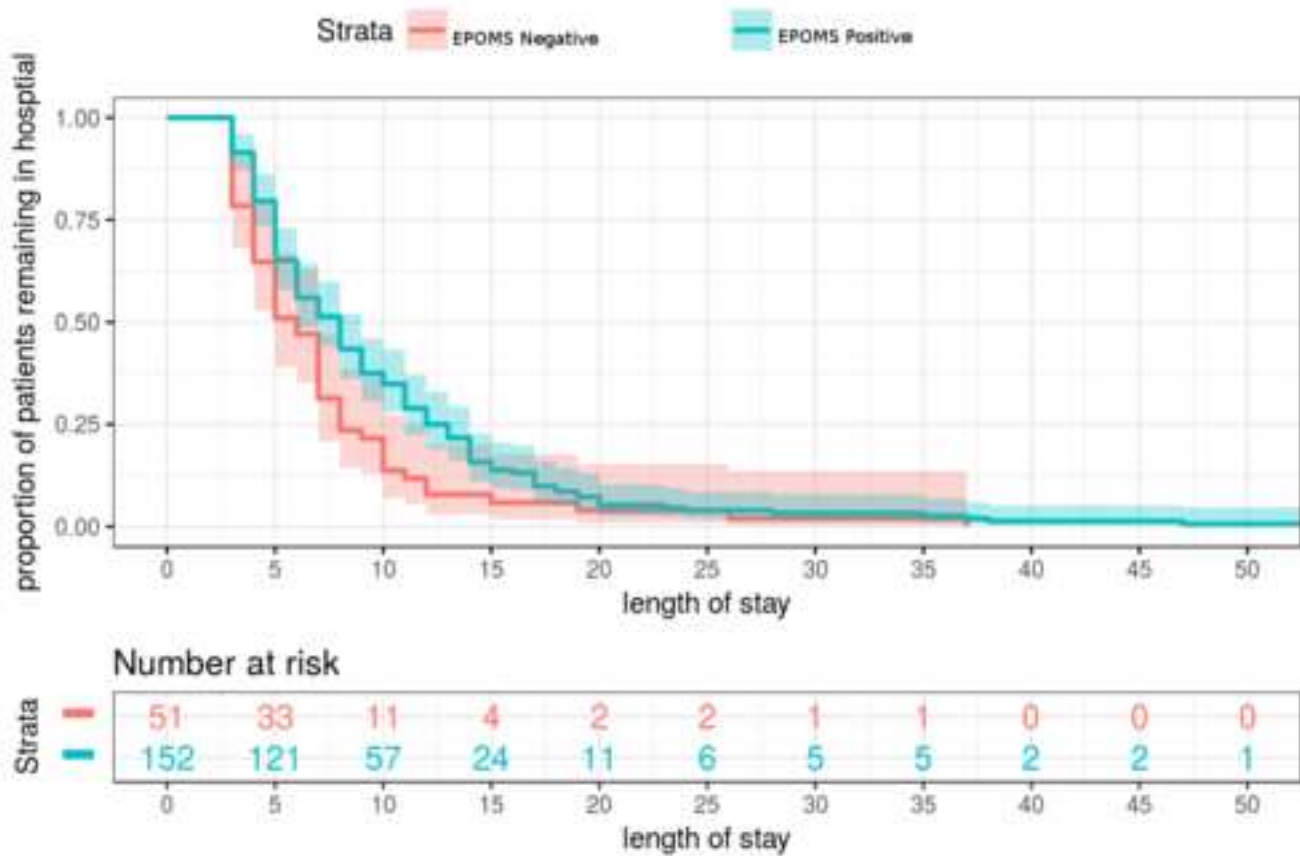
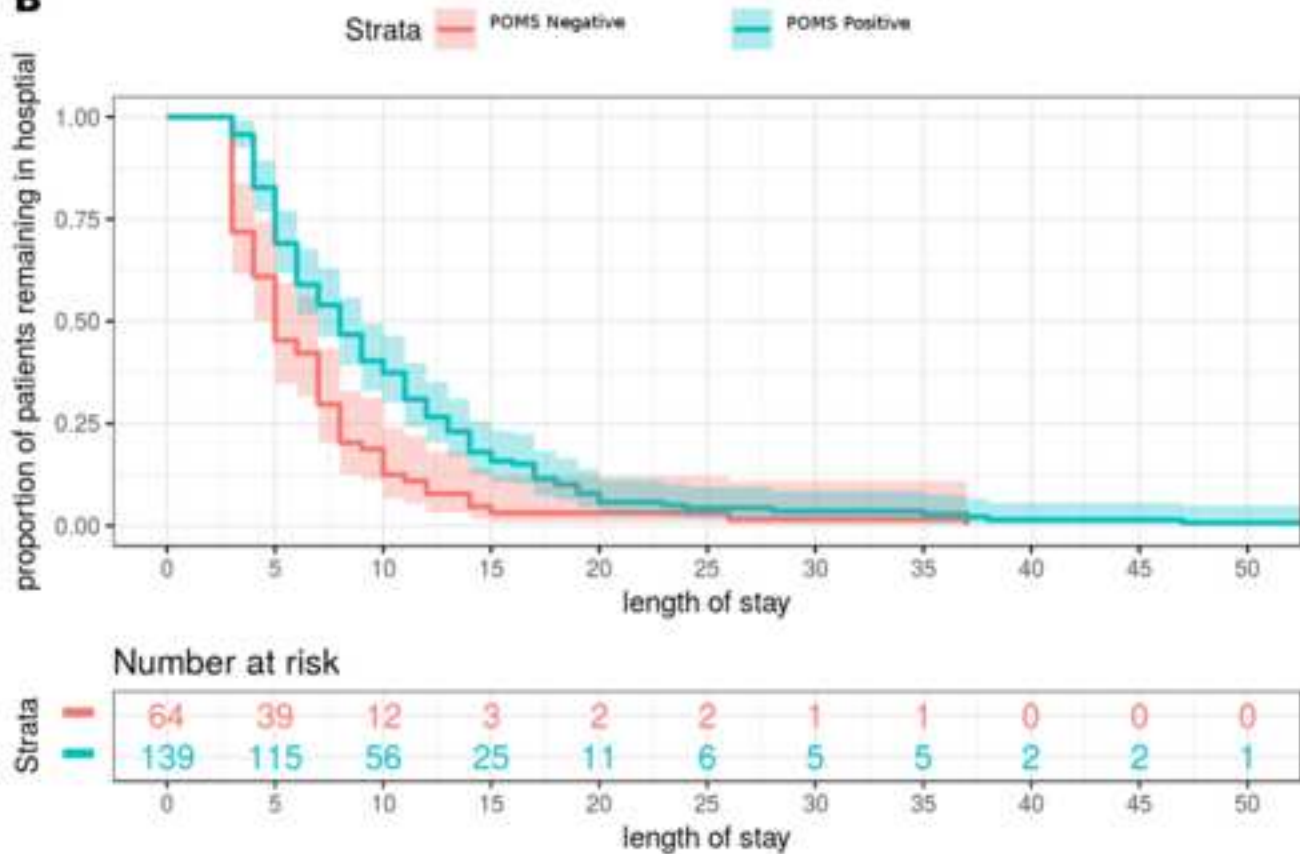
A

Figure 2

[Click here to access/download;Figure;figure2_new.tif](#) 



A**B**

Domain	POMS (n)	Initial EPOMS Score (n)	Final EPOMS Score (n)
Pulmonary	75	75	75
Cardiac	20	101	43
Infectious	79	79	79
Wound	1	1	1
Haematological	11	11	11
Pain	34	34	34
GI	23	20	20
Neurological	18	20	13
Renal	31	48	29

Supplementary Table 1: Difference in incidence of morbidity using POMS (Postoperative Morbidity Score) or our initial electronic markers

	POMS +	POMS -
Initial EPOMS +	138 (TP)	35 (FP)
Initial EPOMS -	1 (FN)	29 (TN)

(A)

	POMS +	POMS -
Final EPOMS +	138 (TP)	14 (FP)
Final EPOMS -	1 (FN)	50 (TN)

(B)

	2 or more POMS +	POMS -
2 or more Initial EPOMS +	70(TP)	47(FP)
< 2 Initial EPOMS	3(FN)	83(TN)

(C)

	2 or more POMS +	<2 POMS
2 or more Final EPOMS +	66 (TP)	19 (FP)
< 2 Final EPOMS	7 (FN)	111 (TN)

(D)

	3 or more POMS +	<3 POMS
3 or more Initial EPOMS +	33 (TP)	26 (FP)
<3 Initial EPOMS	3 (FN)	141 (TN)

(E)

	3 or more POMS +	< 3 POMS
3 or more Final EPOMS +	30 (TP)	10 (FP)
< 3 Final EPOMS	6 (FN)	157 (TN)

(F)

Supplementary Tables 2 A-F: Confusion matrices assessing performance of initial electronic definitions (tables **A**, **C**, **E**) or final EPOMS definitions (tables **B**, **D**, **F**) for the identification of differing degrees of POMS defined morbidity (total POMS score: ≥ 1 , ≥ 2 , ≥ 3). **A-B:** Performance of both definitions when 1 or more domains is positive. **C-D:** when 2 or more domains positive. **E-F:** when 3 or more domains positive. TP = True Positive, FP = False Positive, FN= False Negative, TN = True Negative

Calculations used:

Sensitivity = $TP/(TP+FN)$, Specificity = $(TN/(TN+FP))$, Negative Predictive Value = $(TN/(TN+FN))$, Positive Predictive Value = $(TP/(TP+FP))$