Statistical Graph Interpretation

Prof. Bruce Biccard

UCT Dept of Anaesthesia & Perioperative Medicine
bruce.biccard@uct.ac.za

Introduction

In this tutorial, we will cover the common statistical graphs one encounters in clinical medicine. We will discuss the indications and interpretation of the graphs, as important considerations when appraising the validity (and limitations) of the graph for the presentation of the data. This tutorial is built predominantly around the following works.[1-4]

Graphs used to describe the characteristics of data

1. Histograms

Indication: To show the frequency and the shape of continuous data. Determining whether data is normally distributed (or can be transformed to normality) allows for the use of parametric tests in data analysis.

Interpretation: A histogram allows for a visual inspection of the data distribution. It is possible therefore to determine whether it approaches a bell (or Gaussian) distribution. It shows where there are gaps (no data points), outliers, and whether the data is skewed. The skewness is named by the longer tail i.e. positive skew, has a long right tail.

Limitations: Normality of data cannot be determined by visual inspection of a histogram alone. Other formal statistical tests are needed to describe normality. The reasons for a deviation from normality may be statistically described by the skewness (positive or negative skew) and kurtosis (flat or pointed distribution) of the data.

Data organisation: The data categories are on the X-axis, and the frequency of the data in each category is plotted on the Y-axis.

When you are most likely to use a histogram: Histograms are infrequently presented in peer-reviewed papers. It is most likely that you will use a histogram as part of your exploratory work on your data to determine if it is normally distributed in preparation for the choice of an appropriated statistical test for the analysis your MMed.

Classic example in the anaesthesia literature: One of the original papers describing the anaerobic threshold (AT) in preoperative surgical patients by Paul Older and colleagues presented the following histogram of the distribution of the AT in their surgical patients.[5] In published data it is important to calculate the 95% confidence interval (CI) from the standard deviation (SD) (95% CI=mean $\pm$ 2 SD), to determine if the data presented are actually plausible. In this example, the average oxygen consumption for all patients at AT was 12.4ml/kg/min with a SD=2.7,[5] therefore the 95% CI for the AT was 6 to 17.8ml/kg/min. This is entirely plausible. The data looks normally distributed (with a possible small negative skew), yet we would need to confirm normality with further statistical analysis.

Other complementary graphs of data normality: The Q-Q plot. Here, the observed value (X-axis) is plotted against the expected value (Y-axis). Deviations from a straight line suggest deviation from normality of distribution.

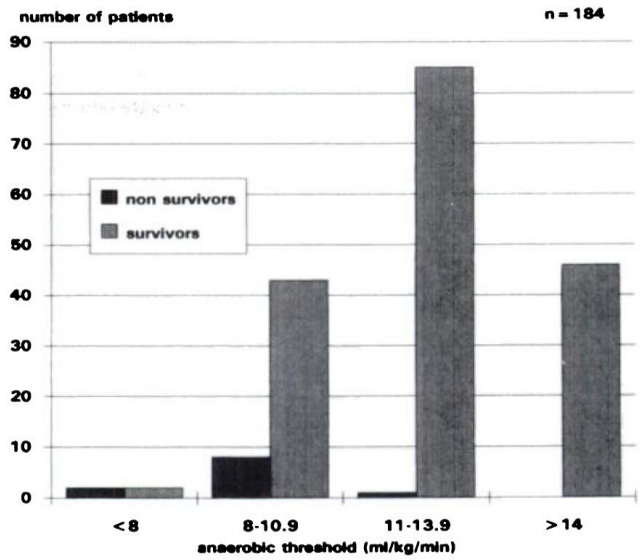

FIGURE 3. Comparison of cardiovascular mortality with anaerobic threshold.

Figure 1. Histogram presented by Paul Older and Colleagues describing the anaerobic threshold in surgical patients[5]

2. Box and whisker plots

Indication: To graphically present the median and interquartile range (IQR) in non-normally distributed data.

Data organisation: The median is presented as the horizontal line in the box, and the length of the box represents the IQR. Extreme values are usually shown when they are more than three box lengths from the upper or lower box end.

When are you most likely to use a box and whisker plot: Presentation of data which is non-normally distributed. This is common with patient biochemical data. The typical presentation is a positive skew due to small proportion of patients having, for example a high serum creatinine, troponin or B-type natriuretic peptide level compared to the population. It is possible to transform non-normally distributed data to a normal distribution in order to use parametric statistical tests. The most common transformation is a log transformation for positively skewed data. This is commonly used with data such as B-type natriuretic peptide levels which are positively skewed.

Classic example in the anaesthesia literature: Cuthbertson and colleagues presented the association between preoperative B-type natriuretic peptides (BNP) and postoperative cardiac events.[6]

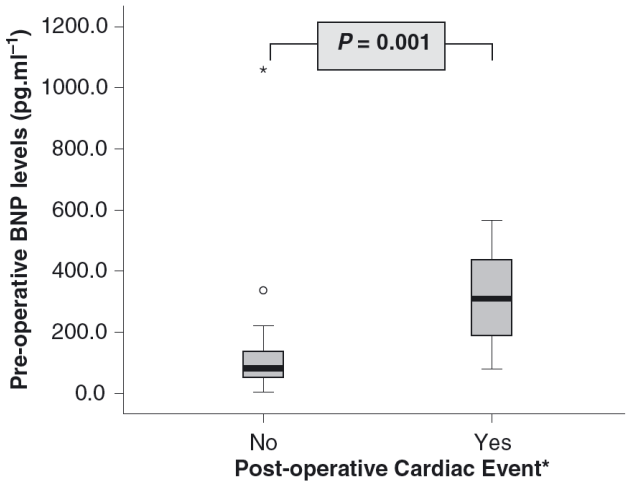

Figure 1 A comparison of pre-operative BNP levels in patients who experienced a postoperative cardiac event with those who did not. Central lines represent medians, boxes 25th and 75th centiles and whiskers represent ranges.

Figure 2. The association between preoperative B-type natriuretic peptides (which are non-normally distributed and positively skewed) and postoperative cardiac events.[6]

Critical appraisal of data normality:

1. Has the normality of the continuous data been determined both visually and statistically?
2. Has the centre and distribution been described appropriately?
3. If the data is claimed to be normally distributed, does the mean±2SD describe a plausible 95% confidence interval?

Graphs used to describe the relationship between two variables

1. Scatter plots or scattergrams

Indication: To understand the nature of the relationship between two continuous variables. This describes how closely two variables are relate, and the amount of variability in one measurement that can be explained by another measurement.

Interpretation: To express the numerical value of the correlation, we calculate the correlation coefficient. The correlation coefficient quantitively expresses both the magnitude (from 0 to 1) and the direction of the correlation (positive i.e. both increase in value together, or negative i.e. one variable deceases, while the other increases). The proportion of variance that can be attributed to one variable, is dependent of the 'coefficient of determination', which is the square of the correlation coefficient e.g. r=0.7, therefore coefficient of determination (r^2) = 0.49 i.e. 49% of the variation can be explained by the relationship between the two variables, and 51% is due to other factors.

Limitations: Firstly, correlation does not represent causation. Causality is dependent of three criteria; i) the causative occurrence needs to precede the effect, ii) covariation of cause and effect i.e. if the cause occurs, then the effect should occur, and iii) elimination of rival causal explanations i.e. if you remove the cause, then the effect should not occur. This is why we use randomised controlled trials to determine causality. Secondly, a lack of correlation using the above tests does not necessary reflect no correlation, as these tests only reflect linear correlation. The two variables may be correlated in another way i.e. some curved relationship.

Data organisation: This is dependent on the type of data. For linear correlations only, the commonly used tests are; i) Pearson's (r) correlation coefficient, which plots two continuous variables (interval or ratio scale) which are normally distributed, ii) Spearman's (rho) correlation coefficient, which plots either two ordinal variables, or one continuous, and one ranked variable, and iii) Kendall's correlation coefficient, which plots two categorical variables.

When are you most likely to use a correlation scattergram: For testing the correlation between tow variables e.g. between two different blood tests measuring the same serum blood marker.
Classic example in anaesthesia literature: In order to determine the association between perioperative volume expansion, and a >15% increase in cardiac output, Yannick le Manach and colleagues look at the association between various haemodynamic variables (change in SBP, DBP, MAP and pulse pressure (PP) variation).[7] The only significant association between volume expansion and an increase in cardiac output of >15% was a > 3% decrease in the PPV i.e PPV after volume expansion (VE) − PPV before VE. The correlation between the change in the PPV and the CO is shown in the figure. The association is negative i.e. the PPV falls, for an increase in the CO. The change in PPV can explain 36% of the change in the CO in this study.

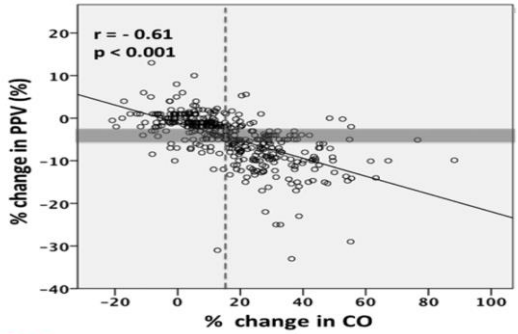

Fig. 2. Relationships between volume expansion–induced changes in PPV and CO. Percent change in PPV = percent changes in pulse pressure variation (defined as pulse pressure variation after volume expansion − pulse pressure variation before volume expansion). Percent change in CO = percent changes in cardiac output induced by volume expansion. The *dashed vertical line* shows the 15% threshold increase in CO defining fluid responsiveness.

Figure 3. The correlation between pulse pressure variation and cardiac output[7]

2. Altman-Bland Plot

Indication: To quantify the agreement between two readings.

Interpretation: Incorrect methods of agreement between two readings are to; i) compare the means (and find no significant difference), or ii) the correlation coefficient (it is a measure of association, not a measure of agreement).

Limitations: This analysis cannot be used to predict one measurement from another. Prediction of a measurement is not adequately addressed by the Altman-Bland method of comparing measurements.
Data organisation: Plot the difference between the methods (A-B) on the Y-axis, against the average of the methods, (A+B)/2, on the X-axis. The mean of the difference (A-B) is the relative bias, and the SD is the estimate of the error.

When are you most likely to use an Altman-Bland plot: Comparing a measurement by a new device/ monitor, against the 'gold standard'.

Classic example in the anaesthesia literature: Suehiro and colleagues comparing 3D transoesophageal echocardiography (TOE) and pulmonary artery (PA) catheter estimation of stroke volume in cardiac surgical patients.[8] They found that the 3D TOE had a negative bias of 1.2ml for the stroke volume with a 95% CI of 11.9ml to -14.3ml).

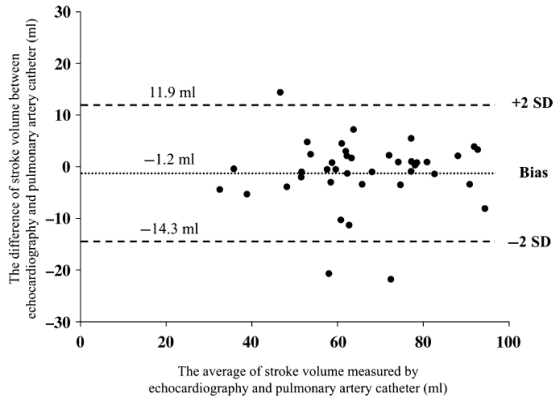

Figure 3 Bland–Altman analysis to investigate the reliability of 3D-transoesophageal echocardiography compared with the pulmonary artery catheter in measuring stroke volume.

Figure 4. Altman-Bland plot of the bias and estimated error of using 3D TOE to determine stroke volume in cardiac surgical patients

Critical appraisal of scatter plots:

1. No matter how small the p-value is, it is the r that is important
2. R is not on a linear scale, but r^2 is
3. A significant p-value, or a high r, do not imply causation.

Receiving operating characteristic (ROC) curves

Indication: To determine a cut-off point in continuously distributed data, that predicts the presence of an outcome. This cut-off point can be either; i) a screening cut point, ii) a diagnostic cut point, or iii) the optimal cut point.

Interpretation: ROC curves are useful for determining a screening cut point i.e. 95% sensitive (curve closest to the top of the Y-axis, where true positives are maximised), or a diagnostic cut point i.e. 95% specific (curve closest to the left of the X-axis, where the rate of false positives is low, and thus true negatives are high).

Limitations: Studies that only present the optimal cut, over-estimate the utility of a test.[9] This is because the data of a single study has been dichotomised to maximise sensitivity and specificity i.e. to determine the point on the ROC curve which is closest to the ideal point i.e. shortest distance to the top left corner of the plot i.e. Distance $=(1\text{-sensitivity})^2 + (1\text{-specificity})^2$

Data organisation: A plot of sensitivity i.e. true positives (on the Y-axis) against 1-specificity i.e. false

positives (on the X-axis). A move upwards shows a true positive result, and a move to the right shows a false positive result. Therefore, the ideal test will be at the top left hand corner of the plot i.e. 100% sensitive and 100% specific. The area under the diagonal line is 0.5 i.e. there is no difference from chance. The success of the test is reported as the area under the curve (AUC).

When are you most likely to use a ROC curve: To determine an appropriate cut point for a test e.g. at what STOP-BANG score should you consider postoperative apnoea a clinical problem.

Classic example in the anaesthesia literature: Choi and colleagues presented the improved discrimination shown by adding preoperative CRP and preoperative BNP to the Revised Cardiac Risk Index (RCRI) in predicting major adverse cardiac events following surgery.[10]

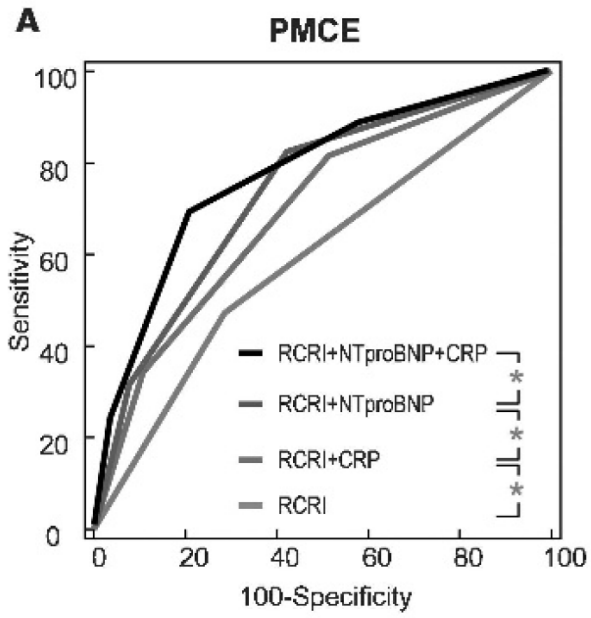

Figure 5. The improvement in discrimination when adding the preoperative CRP or BNP to the Revised Cardiac Risk Index for predicting postoperative major adverse cardiac events.

Critical appraisal of ROC curves:

1. Was the gold standard used to classify the diagnosis?
2. Were the results of the test withheld from the people who classified patients as having the disease (and vice versa)?
3. Are there enough test positive and test negative patients to accurately determine sensitivity and specificity?
4. Are there confidence intervals reported?

Survival plots (or Kaplan-Meier curves)

Indication: To present the time to an outcome in (usually) two different groups (or cohorts). The time to the primary outcome in randomised controlled trials are often presented in survival plots, which are known as Kaplan-Meier curves.

Interpretation: The utility of a survival plot is that it can indicate the time period at which a patient is most likely to be at risk of the outcome i.e. the steepest part of the curve.

Limitations: It is limited by; i) the duration of the follow up i.e. events that occur after the time of completion of follow up are not recorded, or ii) loss to follow up, where patients are censored as per their known last status (which is usually no reported outcome event).

Data organisation: The cumulative outcome is reported on the Y-axis, and the time of the event is

reported on the X-axis. Y-axis data is presented either descending i.e. number of individuals free of the outcome, or ascending i.e. number of individuals who have experienced an outcome event. Each step represents an additional outcome event.

When are you most likely to use a survival plot: To report the time of specific outcomes in two patient cohorts.

Classic example in the anaesthesia literature: The POISE[11] Trial which showed the different beneficial and harmful outcomes associated with perioperative acute beta-blockade, with decreased myocardial infarction and increased stroke and mortality within 30-days of surgery.

Critical appraisal of survival plots:

1. Is the follow up period long enough to identify the outcome?
2. Is there a significant number of patients lost to follow up during the follow up time period?
3. Has the scaling of the Y-axis been manipulated to increase the visual discrimination between groups?

Figure 6. POISE trial outcomes within 30 days of surgery following randomisation to beta-blockade or placebo.[11]

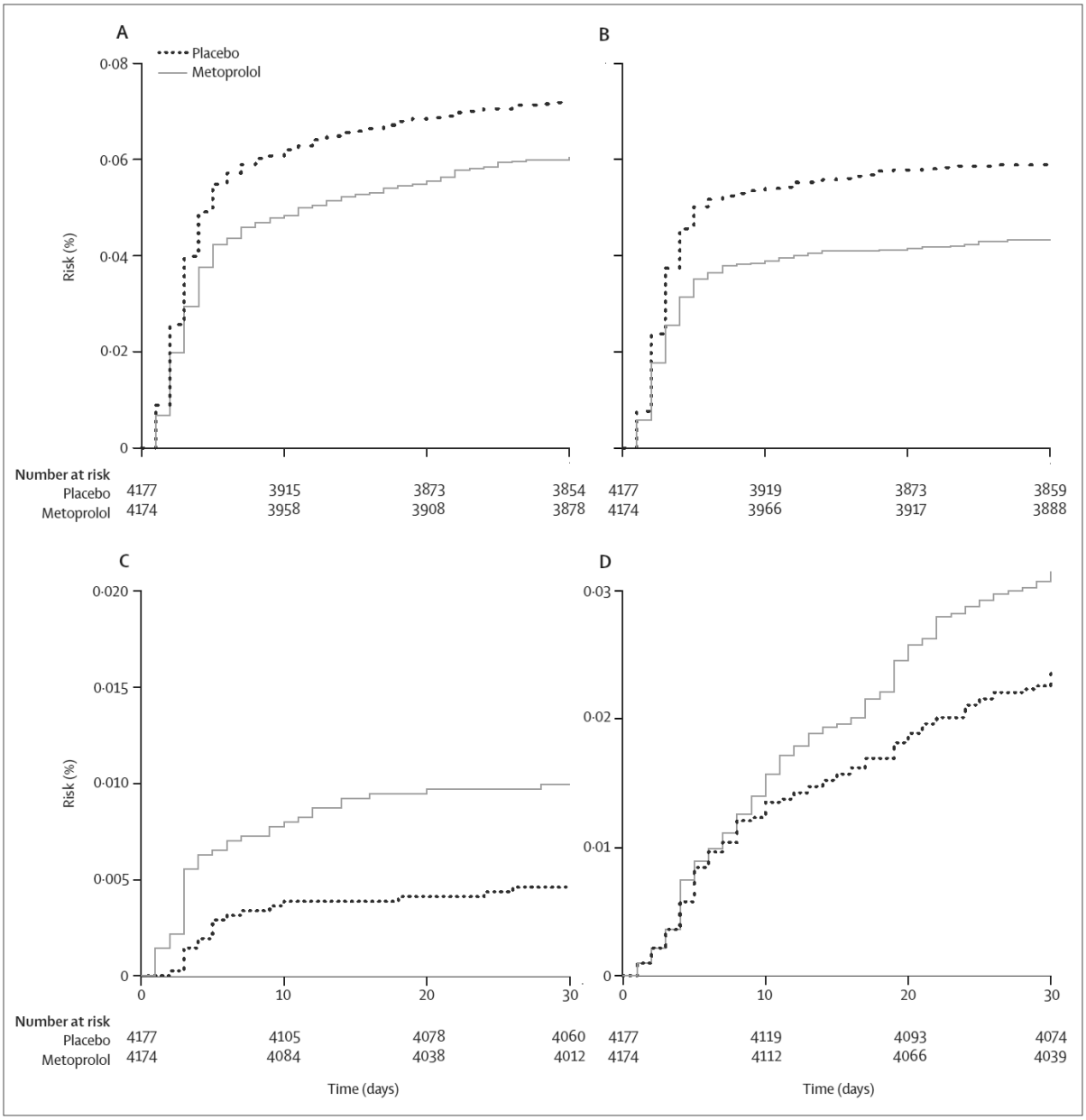

Figure 2: Kaplan-Meier estimates of the primary outcome (A), myocardial infarction (B), stroke (C), and death (D)

Forest plots

Indication: To present the summary data of a meta-analysis.

Interpretation: It summarises the current, aggregated data. If the systematic review was exhaustive, the meta-analysis provides the best evidence of the current data. This information, allows for determination of appropriate, future research projects within the field.

Limitations: Sometimes, there are too few data to allow for a meta-analysis. In this case the data can only be described in a systematic review.

Data organisation: The data for two groups (an intervention group and a control group) are presented. Data can be reported as binary outcomes e.g. alive or dead, or continuous data e.g. change in forced expiratory volume. Continuous data is usually presented as the weighted mean difference (WMD) between groups. The weighted mean difference can be converted back to the standard units of measurement in order to understand the clinical impact of an intervention. The Y-axis lists the studies/ trials included in the meta-analysis. The X-axis shows the line of unity, and the association with the outcome, either as the relative risk (RR), odds ratio (OR), or weighted mean difference (WMD).

When are you most likely to use a forest plot: In a systematic review where there is enough similar data to allow aggregating data from more than one study.

Classic example in the anaesthesia literature: The meta-analysis of perioperative beta-blockade, which conducted subgroup analyses of the secure and insecure studies (following the fraudulent findings against Don Poldermans).[12] The data showed high heterogeneity and treatment effect in the insecure trials, when compared to the secure trials.

Figure 7. A meta-analysis of the effect of acute perioperative beta-blockade on myocardial infarction in secure and insecure clinical trials.[12]

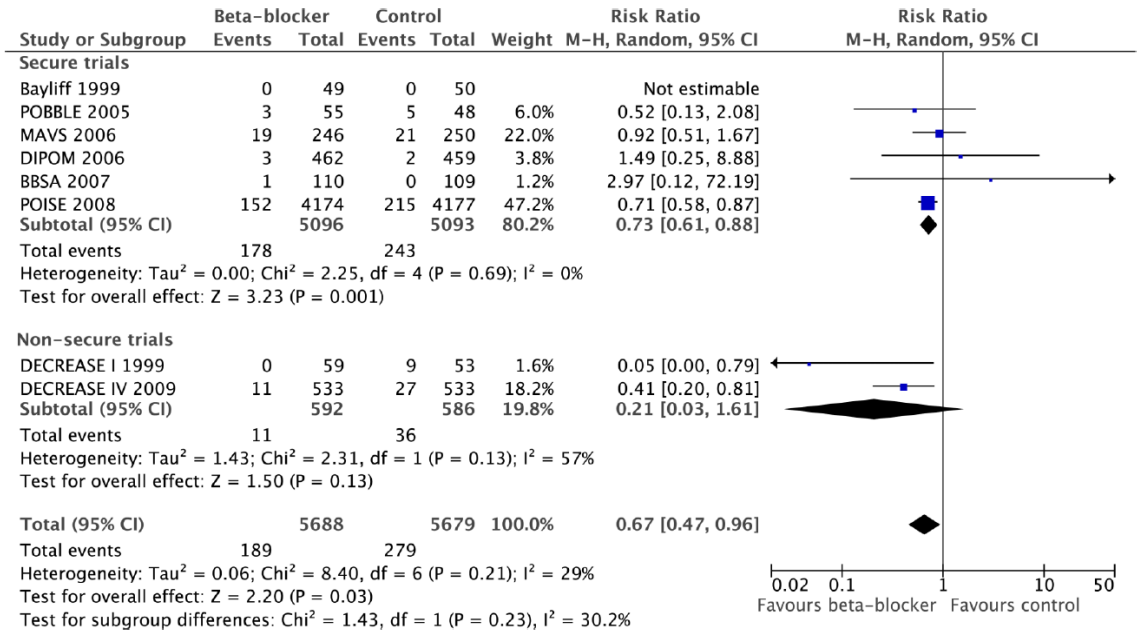

Figure 5 Comparison of effect of perioperative β-blockade on non-fatal myocardial infarction in secure and non-secure trials.

Critical appraisal of forest plots:

1. Is the data similar enough between studies to allow aggregation and meta-analysis? Visual inspection of the point estimates and the CI relative to the line of unity will inform this decision.
2. What is the magnitude of heterogeneity associated with the data, as reflected by the p-value and the I^2 statistic?
3. How precise are the results?

References

1. Peat J, Barton B. Medical statistics. A guide to data analysis and critical appraisal. Great Britain: Blackwell Publishing 2005.
2. Polgar S, Thomas SA. Introduction to research in the health sciences. Third ed. United Kingdom: Churchill Livingstone 1995.
3. Norman GR, Streiner DL. Biostatistics. The bare essentials. Third ed. Connecticut, USA: People's Medical Publishing House 2008.
4. Guyatt GH, Rennie D, Meade MO, et al. Users' guides to the medical literature: A manual for evidence-based clinical practice. Third ed. New York: McGraw-Hill Education 2015.
5. Older P, Smith R, Courtney P, et al. Preoperative evaluation of cardiac failure and ischemia in elderly patients by cardiopulmonary exercise testing. *Chest* 1993;104(3):701-4.
6. Cuthbertson BH, Card G, Croal BL, et al. The utility of B-type natriuretic peptide in predicting postoperative cardiac events and mortality in patients undergoing major emergency non-cardiac surgery. *Anaesthesia* 2007;62(9):875-81. doi: 10.1111/j.1365-2044.2007.05146.x
7. Le Manach Y, Hofer CK, Lehot JJ, et al. Can changes in arterial pressure be used to detect changes in cardiac output during volume expansion in the perioperative period? *Anesthesiology* 2012;117(6):1165-74. doi: 10.1097/ALN.0b013e318275561d [published Online First: 2012/11/09]
8. Suehiro K, Tanaka K, Yamada T, et al. The utility of intra-operative three-dimensional transoesophageal echocardiography for dynamic measurement of stroke volume. *Anaesthesia* 2015;70(2):150-9. doi: 10.1111/anae.12857 [published Online First: 2014/10/01]
9. Potgieter D, Simmers D, Ryan L, et al. N-terminal pro-B-type Natriuretic Peptides' Prognostic Utility Is Overestimated in Meta-analyses Using Study-specific Optimal Diagnostic Thresholds. *Anesthesiology* 2015;123(2):264-71. doi: 10.1097/ALN.0000000000000728
10. Choi JH, Cho DK, Song YB, et al. Preoperative NT-proBNP and CRP predict perioperative major cardiovascular events in non-cardiac surgery. *Heart* 2010;96(1):56-62. doi: 10.1136/hrt.2009.181388
11. Devereaux PJ, Yang H, Yusuf S, et al. Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomised controlled trial. *Lancet* 2008;371(9627):1839-47. doi: 10.1016/S0140-6736(08)60601-7 S0140-6736(08)60601-7 [pii] [published Online First: 2008/05/16]
12. Bouri S, Shun-Shin MJ, Cole GD, et al. Meta-analysis of secure randomised controlled trials of beta-blockade to prevent perioperative death in non-cardiac surgery. *Heart* 2014;100(6):456-64. doi: 10.1136/heartjnl-2013-304262 [published Online First: 2013/08/02]