N-terminal pro-B-type Natriuretic Peptides’ Prognostic Utility Is Overestimated in Meta-analyses Using Study-specific Optimal Diagnostic Thresholds


ABSTRACT

Background: N-terminal fragment B-type natriuretic peptide (NT-proBNP) prognostic utility is commonly determined post hoc by identifying a single optimal discrimination threshold tailored to the individual study population. The authors aimed to determine how using these study-specific post hoc thresholds impacts meta-analysis results.

Methods: The authors conducted a systematic review of studies reporting the ability of preoperative NT-proBNP measurements to predict the composite outcome of all-cause mortality and nonfatal myocardial infarction at 30 days after noncardiac surgery. Individual patient-level data NT-proBNP thresholds were determined using two different methodologies. First, a single combined NT-proBNP threshold was determined for the entire cohort of patients, and a meta-analysis conducted using this single threshold. Second, study-specific thresholds were determined for each individual study, with meta-analysis being conducted using these study-specific thresholds.

Results: The authors obtained individual patient data from 14 studies (N = 2,196). Using a single NT-proBNP cohort threshold, the odds ratio (OR) associated with an increased NT-proBNP measurement was 3.43 (95% CI, 2.08 to 5.64). Using individual study-specific thresholds, the OR associated with an increased NT-proBNP measurement was 6.45 (95% CI, 3.98 to 10.46). In smaller studies (<100 patients) a single cohort threshold was associated with an OR of 5.4 (95% CI, 2.27 to 12.84) as compared with an OR of 14.38 (95% CI, 6.08 to 34.01) for study-specific thresholds.

Conclusions: Post hoc identification of study-specific prognostic biomarker thresholds artificially maximizes biomarker predictive power, resulting in an amplification or overestimation during meta-analysis of these results. This effect is accentuated in small studies. (ANESTHESIOLOGY 2015; 123:264-71)

META-ANALYSES pool results from multiple small studies to determine, as accurately as possible, a true treatment, diagnostic, or prognostic effect. Pooling increases the number of study participants, increases study power, and improves the precision around the estimate of the true effect. Although meta-analytic techniques for pooling data from treatment trials are fairly advanced, applying these to diagnostic and prognostic studies present unique challenge.

In recent times, there has been a surge of interest in the use of biomarkers for disease diagnosis and prognosis. Biomarkers are commonly measured as continuous variables, and in many fields, it has become common practice for authors to dichotomize these continuous biomarker values into high- and low-risk categories. Dichotomization techniques include the use of an optimal discriminatory point threshold to evaluate biomarker prognostic ability, as it is likely that this methodology will overestimate biomarker predictive performance.

What We Already Know about This Topic

• Biomarker prognostic utility is commonly evaluated by identifying a single optimal discrimination threshold for a specific study, often determined post hoc and specifically tailored to the individual study population.

• This study hypothesized that conducting meta-analysis of prognostic studies, where each individual study reported an adjusted odds ratio derived from an optimal study-specific cut-point, would significantly overestimate the prognostic effect of the biomarker—particularly in small studies—as compared with using a single optimal cut-point across all studies. This hypothesis was tested using individual patient data from studies examining the ability of the hormone N-terminal fragment B-type natriuretic peptide to predict the composite outcome of postoperative mortality and nonfatal myocardial infarction at 30 days after noncardiac surgery.

What This Article Tells Us That Is New

• Meta-analysis of studies that made use of a study-specific optimal N-terminal fragment B-type natriuretic peptide threshold resulted in a larger risk point estimate for the prediction of the composite outcome of postoperative mortality and nonfatal myocardial infarction at 30 days after noncardiac surgery compared with using a single threshold across all studies. These data suggest that future biomarker studies should be evaluated as continuous variables rather than making use of post hoc study-specific optimal thresholds, and care should be taken when conducting meta-analysis on studies that have used study-specific optimal thresholds to evaluate biomarker prognostic ability, as it is likely that this methodology will overestimate biomarker predictive performance.

What This Article Tells Us That Is New

• Meta-analysis of studies that made use of a study-specific optimal N-terminal fragment B-type natriuretic peptide threshold resulted in a larger risk point estimate for the prediction of the composite outcome of postoperative mortality and nonfatal myocardial infarction at 30 days after noncardiac surgery compared with using a single threshold across all studies. These data suggest that future biomarker studies should be evaluated as continuous variables rather than making use of post hoc study-specific optimal thresholds, and care should be taken when conducting meta-analysis on studies that have used study-specific optimal thresholds to evaluate biomarker prognostic ability, as it is likely that this methodology will overestimate biomarker predictive performance.

What This Article Tells Us That Is New

• Meta-analysis of studies that made use of a study-specific optimal N-terminal fragment B-type natriuretic peptide threshold resulted in a larger risk point estimate for the prediction of the composite outcome of postoperative mortality and nonfatal myocardial infarction at 30 days after noncardiac surgery compared with using a single threshold across all studies. These data suggest that future biomarker studies should be evaluated as continuous variables rather than making use of post hoc study-specific optimal thresholds, and care should be taken when conducting meta-analysis on studies that have used study-specific optimal thresholds to evaluate biomarker prognostic ability, as it is likely that this methodology will overestimate biomarker predictive performance.

What We Already Know about This Topic

• Biomarker prognostic utility is commonly evaluated by identifying a single optimal discrimination threshold for a specific study, often determined post hoc and specifically tailored to the individual study population.

• This study hypothesized that conducting meta-analysis of prognostic studies, where each individual study reported an adjusted odds ratio derived from an optimal study-specific cut-point, would significantly overestimate the prognostic effect of the biomarker—particularly in small studies—as compared with using a single optimal cut-point across all studies. This hypothesis was tested using individual patient data from studies examining the ability of the hormone N-terminal fragment B-type natriuretic peptide to predict the composite outcome of postoperative mortality and nonfatal myocardial infarction at 30 days after noncardiac surgery.

What This Article Tells Us That Is New

• Meta-analysis of studies that made use of a study-specific optimal N-terminal fragment B-type natriuretic peptide threshold resulted in a larger risk point estimate for the prediction of the composite outcome of postoperative mortality and nonfatal myocardial infarction at 30 days after noncardiac surgery compared with using a single threshold across all studies. These data suggest that future biomarker studies should be evaluated as continuous variables rather than making use of post hoc study-specific optimal thresholds, and care should be taken when conducting meta-analysis on studies that have used study-specific optimal thresholds to evaluate biomarker prognostic ability, as it is likely that this methodology will overestimate biomarker predictive performance.
Materials and Methods

Systematic Review Methodology

We considered studies eligible if they measured NT-proBNP in adult patients 30 days before noncardiac surgery. The primary study outcome was the composite of all-cause mortality and nonfatal myocardial infarction (MI) at 30 days after noncardiac surgery. NT-proBNP is released from the myocardium predominantly in response to myocardial stretch and ischemia, and increases have been associated with adverse postoperative complications. We hypothesized that conducting meta-analysis of prognostic studies, where each individual study reported an OR derived from an optimal study-specific cut-point, would significantly overestimate the prognostic effect of the biomarker—particularly in small studies—as compared with using a single optimal cut-point across all studies. We tested this hypothesis using individual patient data from studies examining the ability of the hormone N-terminal fragment B-type natriuretic peptide (NT-proBNP) to predict the composite outcome of postoperative mortality and nonfatal myocardial infarction (MI) at 30 days after noncardiac surgery. NT-proBNP was only measured postoperatively or where B-type natriuretic peptide (BNP) was measured. Studies that collected relevant data but did not report the outcome of interest were included if the primary outcome could be obtained from the authors.

In July 2013, we searched EMBASE, OVID Health Star, MEDLINE(R) In-Process & Other Non-Indexed Citations and OVID MEDLINE(R), Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews and ProQuest Dissertations, and Theses A&I using the OvidSP search engine (Ovid Technologies, Inc., USA, 2009). We also looked for abstracts from meetings of the American Heart Association and the American Society of Anesthesiologists, our own files, consulted with experts, reviewed reference lists from identified articles, and searched for cited references of key publications in Web of Science. To avoid inclusion of duplicate study data from reports publishing partial results, the study with the most complete follow-up or largest sample size was included. The search terms, including validated prognostic search terms and databases used, are listed in appendix 1.

Two investigators independently screened the titles and abstracts of each citation identified in our search and excluded articles not meeting the study criteria. Citations that the screeners felt had any possibility of meeting eligibility criteria then underwent further full-text review. If either of the two reviewers identified a citation to undergo full review, we obtained the full-text article. Full-text articles were independently evaluated to determine their eligibility for inclusion. Disagreements were solved by consensus, and if this could not be reached, a third adjudicator resolved the matter. Chance-corrected interobserver agreement for study eligibility was tested using kappa statistics.

The authors of eligible studies were then contacted and asked to supply individual patient data for further analysis, that is, preoperative NT-proBNP and primary study outcome at 30 days after surgery. If requested data could not be provided, the study was excluded from the meta-analysis, and similarly, patients from whom these data were missing were also excluded from further analysis.

Statistical Analysis

First, we simulated previously used NT-proBNP biomarker threshold determination and meta-analysis methodology. Using the unadjusted individual patient data, we used ROC curve statistics to determine an optimal NT-proBNP cut-point for each individual study. For each individual study, we then used the study-specific optimal cut-point to dichotomize patients into high and low risk for the primary outcome. We then conducted meta-analysis of the ORs for all the studies, dichotomized according to each study-specific threshold, and derived a pooled OR for the primary outcome.

Second, we combined all patient data, from all studies, creating one single population. Using ROC curve statistics, a single optimal NT-proBNP cut-point for the entire patient population was determined, and the OR associated with
the single optimal cut-point was determined for the entire pooled population.

Third, for each individual study, we used the single population cut-point derived in step 2 to dichotomize patients into high and low risk for the primary outcome. We then calculated an OR for each study and meta-analyzed these results to derive a pooled OR for the primary outcome.

As a subanalysis, we selected those studies with a sample size 100 or less patients, determined study-specific optimal cut-points for each study, and meta-analyzed the dichotomized results. For comparison, we determined a single combined population optimal cut-point, dichotomized each study, and meta-analyzed the dichotomized results.

We used a random-effects model to pool the study results and reported it as a summary OR with its 95% CI. Heterogeneity of included studies was tested with the I² test as well as the chi-square test.

Results
Our database search yielded 1,008 citations. After initial title and abstract screening, 911 citations were excluded. The remaining 97 studies underwent full-text review, and a further 62 studies were excluded for the following reasons: BNP assay used (n = 31), publication retracted or found to be fraudulent (n = 9), cardiac surgery (n = 2), no study endpoints collected (n = 6), editorial/letter to the editor not presenting original data (n = 3), meta-analysis (n = 1), and nonsurgical (n = 4). We identified 35 eligible studies, representing 22 unique patient cohorts, of which 8 were excluded, as we were unable to obtain individual patient data on them. Fourteen studies (n = 2,196) were used in our analysis (fig. 1).

In the first meta-analysis, using study-specific thresholds for each individual study—shown in appendix 2—the OR associated with the primary outcome was 6.45 (95% CI, 3.98 to 10.46; I² = 45%; fig. 2).

In the second meta-analysis, using a single cohort NT-proBNP threshold of 367.15 pg/ml, the OR associated with the primary outcome was 4.38 (95% CI, 3.31 to 5.81; appendix 3). In the final meta-analysis, where the single cohort threshold of 367.15 pg/ml was used to determine an OR for each of the individual studies, the OR associated with the primary outcome was 3.43 (95% CI, 2.08 to 5.64; I² = 39%; fig. 3).

The subanalysis of studies with 100 or less patients included seven studies. Using individual study-specific thresholds resulted in an OR of 14.38 (95% CI, 6.08 to 34.01; I² = 0%), whereas using the single cohort threshold of 367.15 pg/ml yielded an OR of 5.4 (95% CI, 2.27 to 12.84; I² = 0%; figs. 4 and 5, respectively).

Discussion
Statement of Principle Findings
Meta-analysis of studies that made use of a study-specific optimal NT-proBNP threshold resulted in a larger risk point estimate for the prediction of the composite outcome of postoperative mortality and nonfatal MI at 30 days after noncardiac surgery (OR, 6.45; 95% CI, 3.98 to 10.46) compared with using a single threshold across all studies (OR, 3.43; 95% CI, 2.08 to 5.64). This effect was more pronounced in studies with 100 or less patients, where meta-analysis of study-specific thresholds resulted in an OR of 14.38 (95% CI, 6.08 to 34.01) as compared with an OR of 5.4 (95% CI, 2.27 to 12.84) when a single threshold was used for all studies.

Interpretation
This overestimation or amplification effect that we have demonstrated can be attributed to the methodology by which study-specific optimal prognostic thresholds were determined in a post hoc manner. This methodology artificially maximizes the predictive power of the biomarker within the individual study. This effect can be appreciated in the current literature by reviewing ORs reported in meta-analysis that have examined the risk of postoperative mortality and cardiovascular complications associated with preoperative natriuretic peptide increases. Early meta-analyses published between 2008 and 2009 reported ORs of 17.37 (95% CI, 3.31 to 91.15), 4 44.2 (95% CI, 7.6 to 257.0), 3 and 19.77 (95% CI, 13.18 to 29.65). 7 The majority of the individual studies included in these meta-analyses made use of a study-specific optimal threshold. In contrast, a recent individual patient-level data meta-analysis that made use of a single cohort threshold for all studies reported increased preoperative natriuretic peptide measurements to be associated with an OR 3.40 (95% CI, 2.57 to 4.47) for the outcome of mortality and nonfatal MI. 22 We hypothesize that this substantially lower OR ratio may in part be due to the amplification effect described in this article.

Natriuretic peptide measurements are a continuous data variable. Continuous variables are seen to have limited clinical utility and so ROC curves are often used to determine a single optimal cut-point to dichotomize patients into high and low preoperative risk groups. The ROC optimal cut-point is determined by optimizing the rate of true positives while minimizing the rate of false positives to determine the single value reflecting the highest accuracy for the outcome of interest. 1 The thresholds identified using this methodology provide a study-specific threshold that optimally discriminates high-risk patients from low-risk patients within that specific patient population. In preoperative natriuretic peptide prognostic studies, these thresholds vary dramatically (e.g., BNP; 35, 50, 108.5, and 165 pg/ml). 23

The risks inherent in the dichotomization of continuous variables have been extensively highlighted. 24,25 Although dichotomization simplifies statistical analysis and interpretation, and improves clinical applicability of the results, this comes at the cost of information loss. This is even more of a problem when the study is small in size. Furthermore, dichotomization may also increase the risk of a positive
result being a false positive as individuals close to, but on opposite sides of the determined threshold, are characterized as being very different rather than very similar. Only when a threshold effect value truly exists, is dichotomization of data appropriate. That is, if we can assume some binary split of the continuous variable, will create two relatively distinct but homogeneous groups with respect to a particular outcome.

**Implication for Future Research**

Consideration of statistical power in studies examining diagnostic performance is often overlooked. We have demonstrated that the overestimation effect is considerably more pronounced when small studies are analyzed. To minimize this, a sample size calculation should be undertaken to ensure that prognostic studies are large enough to provide robust results.

The prognostic ability of biomarkers such as NT-proBNP should be evaluated as a continuous variable (log-transformed if appropriate) within regression models. Where the function form of the continuous variable is not known, spline or multivariable fractional polynomial modeling should be used. Finally, investigators should define exploratory thresholds *a priori*, rather than making use of *post hoc* determined study-specific optimal thresholds.

An alternative method to providing a single threshold that dichotomizes the population is to provide two thresholds, separated by a “gray zone.” The first cutoff is chosen to include the diagnosis with near-certainty, whereas the second is chosen to exclude the diagnosis with near-certainty. The two cutoffs and gray zone comprise three biomarker intervals that can be associated with their respective likelihood ratios. The positive likelihood ratio of the highest value of the biomarker in the gray zone is considered to include the diagnosis and the negative likelihood ratio (LHR) of the lowest value to exclude the diagnosis. This is often called the “interval LHR” and results in less loss of information and less distortion than choosing a single cutoff. This methodology may however...
still prove problematic and lead to possible overestimation, as the cutoff values are determined post hoc.

Care should be taken when conducting meta-analysis on studies that have used study-specific optimal thresholds to evaluate biomarker prognostic ability, as it is likely that this methodology will overestimate the biomarker predictive performance. Individual patient data meta-analysis may address some of these limitations.

**Limitations**

This analysis was limited by our inability to obtain data from all eligible studies. However, we believe that the large number of patients who were included in our analysis suggests that our findings can be widely generalized. We have described this phenomenon using data from NT-proBNP studies, and it is possible that a similar effect may not be seen with other biomarkers, where more natural thresholds may exist.
Conclusion

Meta-analysis of studies that made use of a study-specific optimal NT-proBNP threshold resulted in a larger risk point estimate for the prediction of the composite outcome of postoperative mortality and nonfatal MI at 30 days after noncardiac surgery (OR, 6.45; 95% CI, 3.98 to 10.46) compared with using a single threshold across all studies (OR, 3.43; 95% CI, 2.08 to 5.64). This effect was more pronounced in studies with 100 or less patients. Future biomarker studies should be evaluated as continuous variables rather than making use of post hoc study-specific optimal thresholds, and care should be taken when conducting meta-analysis on studies that have used study-specific optimal thresholds to evaluate biomarker prognostic ability, as it is likely that this methodology will overestimate biomarker predictive performance.

Acknowledgments

Support was provided solely from institutional and/or departmental sources.

Competing Interests

Dr. Mahla has spoken for and received consulting fees from CSL Behring Biotherapies for Life (Vienna, Austria), AstraZeneca (Vienna, Austria), and Boehringer Ingelheim (Ingelheim, Germany). She has received study grants from CSL Behring Biotherapies for Life and Novo Nordisk Pharma GmbH (Vienna, Austria). The other authors declare no competing interests.

Correspondence

Address correspondence to Dr. Rodseth: Department of Anaesthesiology, Grey’s Hospital, Private Bag X9001, Pietermaritzburg.
Appendix 1. Example of Search Conducted on MEDLINE

<table>
<thead>
<tr>
<th>Search Terms</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Natriuretic peptide OR natriureti*).mp.</td>
<td>90,064</td>
</tr>
<tr>
<td>2 (BNP OR B type natriureti* OR B-type natriureti* OR Brain natriureti*).mp.</td>
<td>42,364</td>
</tr>
<tr>
<td>3 (NT-pro BNP OR NT-proBNP OR NT-pro-BNP OR N terminal proBNP OR N terminal pro-BNP OR N-terminal pro-BNP OR N-terminal pro-brain natriureti* OR N-terminal pro-B-type natriureti*) OR N-terminal pro-B type natriureti*).mp</td>
<td>13,484</td>
</tr>
<tr>
<td>4 (Surgery OR operative OR noncardiac).mp.</td>
<td>3,308,271</td>
</tr>
<tr>
<td>5 1 or 2 or 3</td>
<td>94,975</td>
</tr>
<tr>
<td>6 4 and 5</td>
<td>4,456</td>
</tr>
<tr>
<td>7 Prognosis.sh. or diagnosed.tw. or cohort:.mp. or predictor:.tw. or death.tw. or exp models, statistical/</td>
<td>4,837,473</td>
</tr>
<tr>
<td>8 6 and 7</td>
<td>1,433</td>
</tr>
<tr>
<td>9 Remove duplicates from 8</td>
<td>876</td>
</tr>
</tbody>
</table>

No additional search filters were used
BNP = B-type natriuretic peptide; NT-proBNP = N-terminal B-type natriuretic peptide.

Appendix 2. Study-specific NT-proBNP Cut-points Determined Using ROC Statistics

<table>
<thead>
<tr>
<th>Study</th>
<th>Threshold</th>
<th>AROC</th>
<th>Low CI</th>
<th>Upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardinale et al.</td>
<td>8</td>
<td>130</td>
<td>0.647</td>
<td>0.518</td>
</tr>
<tr>
<td>Chong et al.</td>
<td>13</td>
<td>1,908</td>
<td>0.73</td>
<td>0.546</td>
</tr>
<tr>
<td>Chong et al.</td>
<td>18</td>
<td>824</td>
<td>0.708</td>
<td>0.54</td>
</tr>
<tr>
<td>Chong et al.</td>
<td>13</td>
<td>1,306.5</td>
<td>0.842</td>
<td>0.71</td>
</tr>
<tr>
<td>Manikandan et al.</td>
<td>21</td>
<td>27.5</td>
<td>0.843</td>
<td>0.681</td>
</tr>
<tr>
<td>Waliszek et al.</td>
<td>17</td>
<td>243.2</td>
<td>0.686</td>
<td>0.499</td>
</tr>
<tr>
<td>Lurati Buse et al.</td>
<td>20</td>
<td>368.5</td>
<td>0.652</td>
<td>0.59</td>
</tr>
<tr>
<td>Schutt et al.</td>
<td>11</td>
<td>798.5</td>
<td>0.744</td>
<td>0.581</td>
</tr>
<tr>
<td>Mahla et al.</td>
<td>9</td>
<td>277.5</td>
<td>0.71</td>
<td>0.591</td>
</tr>
<tr>
<td>Cnotliwy et al.</td>
<td>14</td>
<td>1,098</td>
<td>0.813</td>
<td>0.62</td>
</tr>
<tr>
<td>Rajagopalan et al.</td>
<td>16</td>
<td>311</td>
<td>0.7</td>
<td>0.582</td>
</tr>
<tr>
<td>Oscarsson et al.</td>
<td>10</td>
<td>2,076.5</td>
<td>0.61</td>
<td>0.5</td>
</tr>
<tr>
<td>Farzi et al.</td>
<td>15</td>
<td>1,556.5</td>
<td>0.803</td>
<td>0.733</td>
</tr>
<tr>
<td>Bogicevic et al.</td>
<td>12</td>
<td>409.7</td>
<td>0.863</td>
<td>0.772</td>
</tr>
</tbody>
</table>

AROC = area under receiver operator curve; NT-proBNP = N-terminal B-type natriuretic peptide; ROC = receiver operator curve.

Appendix 3: Meta-analysis Using a Single Cohort NT-proBNP Threshold of 367.15 pg/ml to Determine the OR Associated with the Primary Outcome

M-H = Mantel-Haenszel; NT-proBNP = N-terminal fragment of pro-B-type natriuretic peptide; OR = odds ratio.

References

9. Mahla e, Baumann A, Rehak P, Watzinger N, Vicenzi MN.
5. Rodseth RN, Padayachee I, Biccard BM: A meta-analysis of the
4. Devereaux PJ: is a pre-operative brain natriuretic peptide or
20. Lurati Buse GA, Leibowitz DW, Biccard BM: The usefulness of terminal pro-brain natriuretic peptide to predict postoperative cardiac complications and long-term mortality after emergency lower limb orthopedic surgery. Am J Cardiol 2010; 106:865–72
5. Rodseth RN, Padayachee I, Biccard BM: A meta-analysis of the
4. Devereaux PJ: is a pre-operative brain natriuretic peptide or