



Published in final edited form as:

Clin Trials. 2017 December ; 14(6): 629–638. doi:10.1177/1740774517723588.

Evaluating biomarkers for prognostic enrichment of clinical trials

Kathleen F Kerr¹, Jeremy Roth¹, Kehao Zhu¹, Heather Thiessen-Philbrook², Allison Meisner¹, Francis Perry Wilson², Steven Coca³, and Chirag R Parikh^{2,4}

¹Department of Biostatistics, University of Washington, Seattle, WA, USA

²Program of Applied Translational Research, Yale University School of Medicine and VA Medical Center, New Haven, CT, USA

³Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁴Section of Nephrology, Yale University School of Medicine, New Haven, CT, USA

Abstract

Background/Aims—A potential use of biomarkers is to assist in prognostic enrichment of clinical trials, where only patients at relatively higher risk for an outcome of interest are eligible for the trial. We investigated methods for evaluating biomarkers for prognostic enrichment.

Methods—We identified five key considerations when considering a biomarker and a screening threshold for prognostic enrichment: (1) clinical trial sample size; (2) calendar time to enroll the trial; (3) total patient screening costs and the total per-patient trial costs; (4) generalizability of trial results; (5) ethical evaluation of trial-eligibility criteria. Items (1)–(3) are amenable to quantitative analysis. We developed the Biomarker Prognostic Enrichment Tool (BioPET) for evaluating biomarkers for prognostic enrichment at varying levels of screening stringency.

Results—We demonstrate that both modestly prognostic and strongly prognostic biomarkers can improve trial metrics using BioPET. BioPET is available as a webtool at <http://prognosticenrichment.com> and as a package for the R statistical computing platform.

Conclusion—In some clinical settings, even biomarkers with modest prognostic performance can be useful for prognostic enrichment. In addition to the quantitative analysis provided by BioPET, investigators must consider the generalizability of trial results and evaluate the ethics of trial eligibility criteria.

Keywords

Biomarker; risk prediction; clinical trial; prognostic enrichment

Background/Aims

Catalyzed by advances in modern molecular technologies, there has been a surge of interest in biomarkers for multiple purposes, including early detection of disease,^{1, 2} improved diagnosis,^{3, 4} and optimizing treatment.^{5, 6} One popular use of biomarkers is for the “prognostic enrichment” of clinical trials.⁷⁻⁹ For an intervention intended to reduce the occurrence of some unwanted clinical event, a prognostically enriched trial enrolls only patients at relatively higher risk of experiencing the event without the intervention. The rate of the clinical event will be higher in the “enriched” study population, which means that a smaller sample size can be used in the trial while maintaining adequate power to detect a treatment effect.¹⁰ By enabling smaller trials, prognostic enrichment can produce greater efficiency in evaluating new interventions, with potential benefits for patients, sponsors, and public health.

To clarify the scope of this article, we contrast prognostic and predictive enrichment. Prognostic enrichment, the topic of this article, increases the absolute risk of the clinical event in the study population. As described in the Food and Drug Administration draft guidance,⁸ prognostic enrichment means selecting patients with greater likelihood of having the clinical event. With a prognostic enrichment strategy, the biomarker is not expected to predict treatment efficacy. On the other hand, predictive enrichment is a strategy of selecting patients more likely to respond to the treatment. For example, some cancer treatments are only expected to be effective if the cancer cells express certain proteins. Markers of those proteins would be considered predictive biomarkers and a trial of the treatment would likely employ a predictive enrichment strategy. We do not consider predictive enrichment further in this article.

While there is a substantial literature on predictive biomarkers, little has been written about how to evaluate a biomarker when prognostic enrichment is its intended use. This article discusses how to evaluate a biomarker for its prognostic enrichment potential and describes software for this purpose called BioPET (Biomarker Prognostic Enrichment Tool). In this article, the term “biomarker” can refer to either a single measured characteristic or a “composite biomarker”⁸ combining multiple biomarkers or other predictors.

The motivation for prognostic enrichment

Conducting trials in enriched study populations has been common for cardiovascular outcomes.^{7, 8} The principles discussed in this article are broad; examples come from nephrology because that is the specialty of several coauthors. As a motivating example, consider the population of patients with autosomal dominant polycystic kidney disease. Suppose that a novel therapy has been developed that may improve outcomes, and that a trial is designed where the primary endpoint is a substantial decline in renal function (defined as a 30% worsening in glomerular filtration rate relative to a patient’s baseline). In an un-enriched cohort, such a decline occurs for about 20% of these patients in a three-year period.¹¹ Evaluating the novel therapeutic would require a large clinical trial. For example, to have 90% power to detect a relative 30% reduction in this endpoint would require a trial with 1643 patients.

Next, suppose that a biomarker has some ability to distinguish patients with autosomal dominant polycystic kidney disease at greater risk of decline. For example, perhaps 40% of biomarker-positive patients experience substantial renal function decline, compared to 20% of patients overall. With a larger event rate in the “enriched” group, the number of patients needed to have 90% power – 651 patients – is smaller than the trial that enrolls patients regardless of risk. This might mean that a previously-prohibitively expensive trial becomes feasible.⁷⁻⁹

Prognostic enrichment also has potential ethical advantages. An enriched trial selects against patients who are unlikely to experience the event and thus unlikely to benefit from the intervention. Ethical considerations may have partly motivated prognostic enrichment in trials of tamoxifen for preventing breast cancer;¹² tamoxifen can have serious side effects so tamoxifen treatment is not justified in women at very low risk of breast cancer. We revisit a consideration of ethics in the next section.

Biomarkers for prognostic enrichment

The most common metric for summarizing a biomarker’s performance is the area under the receiver operating characteristic curve. However, the area under the receiver operating characteristic curve does not translate to any measure of clinical or public health impact.¹³⁻¹⁵ In particular, this metric does not describe the impact of using a biomarker for prognostic enrichment.

Prognostic of enrichment clinical trials: Key considerations: The setting is a trial for an intervention intended to reduce the occurrence of a clinical event in a patient population. The goal is to establish whether the intervention is useful for the population, or for a subset of the population at highest risk of the event. For example, early trials of statins included only patients at higher risk of cardiovascular events.^{16, 17}

There are five key aspects to consider when contemplating a prognostic enrichment strategy: (1) Clinical trial sample size; (2) Calendar time to enroll the trial; (3) Total patient screening costs and the total of per-patient costs for patients in the trial; (4) Generalizability of trial results; and (5) Ethical evaluation of trial-eligibility criteria. Reducing trial sample size (1) may be the primary motivation for considering prognostic enrichment. With more stringent levels of patient screening – that is, greater levels of enrichment – a well-powered trial can use a smaller sample size. However, depending on the performance of the biomarker, the calendar time (2) required to enroll a prognostically enriched trial can be either longer or shorter than the non-enriched trial. We will show examples of both situations (Figure 1). If the biomarker is only weakly prognostic of the outcome, then the prognostic enrichment strategy usually increases the calendar time needed to enroll the trial. Cost (3) is another important consideration. An enrichment strategy can be appealing because smaller trials are less expensive. However, savings in trial costs can be offset by additional costs of biomarker measurement that accompany a prognostic enrichment strategy, although typically the costs for measuring biomarkers are many times smaller than the cost of including a patient in a trial.

By definition, any clinical trial does not test the intervention on patients who do not meet the trial's eligibility criteria. For a prognostically-enriched trial that demonstrates treatment efficacy, there will naturally be questions about whether the treatment effect extends to patients at risk for the clinical event but not eligible for the trial. In particular, patients who were just below the biomarker eligibility threshold are excluded because of the need to keep the trial sample size and costs manageable, not because those patients were expected not to benefit from the treatment, if effective. Therefore, if a prognostically enriched trial demonstrates treatment efficacy, a new trial might be called for with broader eligibility criteria.^{8, 18} On the other hand, if the prognostically enriched trial is well powered and does not demonstrate treatment efficacy, then there would likely be no reason to consider additional trials in the patient population who are at lower risk of the clinical event. These considerations of generalizability -- item (4) on the list above -- are not amenable to quantitative analysis, but are important and should not be overlooked.

Similarly, the ethics of the trial-eligibility criteria (5) must always be considered. For treatments with substantial toxicities or serious potential side effects, it is only ethical to conduct a trial among patients for whom the potential benefits justify the risks. In the setting of a prognostic biomarker, such ethical considerations might favor restricting the trial to patients with sufficiently poor prognoses, i.e. a higher level of prognostic enrichment. In addition, using a prognostic enrichment strategy to reduce the trial sample size might be judged ethically favorable because a smaller trial exposes fewer patients to the unknown harms of an investigational intervention. Of course, the downside to stricter eligibility criteria is that patients for whom the treatment may be sufficiently effective are excluded.

Methods

Evaluating a biomarker for prognostic enrichment requires specifying the context in several areas: (1) **Clinical context**, specifically the rate of the clinical endpoint in the non-intervention group. (2) **Statistical testing specifications** for the primary trial hypothesis, including the α -level and power. (3) **Biomarker performance**. The user provides the following inputs to BioPET:

1. The **event rate in the non-intervention group** without enrichment. For example, 20% of patients in the non-intervention group are expected to experience the clinical event.
2. **Treatment efficacy**. This is the effect size of the intervention and is represented by the percent reduction in the event rate for patients receiving the intervention that the trial should be powered to detect. For example, one may want to design a trial powered to detect a 30% reduction in the event rate.
3. **Statistical testing parameters**. The user specifies whether the study will use one-sided or two-sided statistical hypothesis testing and the α -level (type I error rate) of the test. For example, we might plan for one-sided testing of the null hypothesis and using $\alpha = 0.025$.
4. **Power**. For example, we might design our clinical trial to have 90% power to detect the specified treatment effect.

5. **Prognostic capacity of biomarker or risk model**, as summarized by the general shape of the receiver operating characteristic curve and the area under the curve. For example, we may have a biomarker with area under the curve of 0.7 and would like to explore whether its performance is sufficient to be useful for prognostic enrichment. Because an area under the curve cannot fully characterize the prognostic capacity of a biomarker, BioPET allows investigators to choose from three prototypical receiver operating characteristic curves with different shapes that all share the user-specified area under the curve. If, for example, an investigator knows that the sensitivity of the biomarker increases rapidly as the false positive rate increases, s/he will choose the prototypical receiver operating characteristic curve that matches this.
6. (Optional input) The **costs** of (a) enrolling and retaining a patient in a trial and (b) the cost of screening a patient for eligibility using the biomarker. For example, the cost of enrolling and retaining a patient in a trial may be \$10,000 and the cost of measuring the biomarker to determine patient eligibility for the trial is \$1000. When these costs are provided, BioPET calculates the total overall per-patient cost for different enrichment strategies.

BioPET produces results using the following methodology.

Biomarker distributions

BioPET uses parametric models for the biomarker, one producing a symmetric receiver operating characteristic curve and two others producing “shifted” receiver operating characteristic curves. For the symmetric curve, biomarker data are simulated as $N(0,1)$ in non-event patients and $N(K, 1)$ in event patients, where K satisfies the relation $\text{area under the curve} = \Phi(K/2)$ and $\Phi(\cdot)$ is the cumulative distribution function of a standard Normal random variable.

For the “shifted” receiver operating characteristic curves, BioPET uses lomax distributions for biomarker data to preserve concavity,¹⁹ since biNormal receiver operating characteristic curves are not concave except in the case of symmetry. For the left-shifted receiver operating characteristic curve, BioPET simulates biomarker data for non-event patients from a lomax distribution with scale and shape parameters set to 1. For event patients, BioPET simulates biomarker data from a lomax distribution with scale parameter 1 and shape parameter set to $(1-\text{area})/\text{area}$, where “area” means the area under the receiver operating characteristic curve.

For the right-shifted receiver operating characteristic curve, BioPET simulates data from a lomax distribution for non-event patients with shape parameter set to $(1-\text{area})/\text{area}$ and scale parameter 1 (“area” means the area under the receiver operating characteristic curve). For event patients BioPET simulates biomarker data from a lomax distribution with scale and shape parameters set to 1. BioPET then used the negative values of these simulated data as the “biomarker,” which produces the right-shifted receiver operating characteristic curve.

Event rate among biomarker-positive patients

First, BioPET simulates data for 500,000 hypothetical patients as described above using the user-supplied event rate to determine the proportions of event and non-event patients. For each screening threshold, the proportion of event patients among all patients exceeding the threshold determines the event rate among biomarker-positive patients.

Sample size

Based on the desired power $0 < 1 - \beta < 1$, Type I error rate $0 < \alpha < 1$, event rate without intervention $0 < \pi < 1$, and event rate with intervention $0 < \tau < 1$, the sample size across the two arms of the trial for a two-sided test is ²⁰

$$\text{SampleSize} = 2 \times \frac{\left(\phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{2 \left(\frac{\pi + \tau}{2}\right) \left(1 - \frac{\pi + \tau}{2}\right)} + \phi^{-1}(1 - \beta) \sqrt{\pi(1 - \pi) + \tau(1 - \tau)} \right)^2}{(\pi - \tau)^2},$$

where π , τ , and $\phi^{-1}(x)$ is the quantile function of the standard Normal distribution such that $\phi^{-1}(x) = z$ where $P[Z < z] = x$. For a one-sided test, the formula is the same except that $\phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ is replaced with $\phi^{-1}(1 - \alpha)$

Total screened

The screening threshold p implies that $1/(1-p)$ patients must be screened to identify one patient eligible for the trial. Therefore, Total patients screened = Trial Sample Size $/ (1-p)$.

Total cost—Let $C1$ be the cost of enrolling a patient in a clinical trial and $C2$ be the cost of screening a patient for a trial using the biomarker. The trial Sample Size is denoted SS . We assume that only patients who agreed to participate in the trial will be screened. The total cost of the trial with screening threshold $p > 0$ is

$C1 \cdot \text{SampleSize} + C2 \cdot \frac{\text{SampleSize}}{(1-p)} = \text{SampleSize} (C1 + C2 / (1-p))$. When $p=0$, there is no biomarker-based screening so total cost = $C1 \cdot \text{SampleSize}$ when $p=0$. The percent reduction in total cost is calculated relative to an unenriched trial (no biomarker screening).

Biomarker combinations and bootstrap intervals (R implementation only)

When multiple biomarkers are specified, they are combined into a single linear combination using logistic regression. The model is fit using the data provided and then fixed. Uncertainty in results is expressed via 95% bootstrap intervals, calculated as follows. N bootstrap samples (default is 1000 bootstrap samples) of the data are simulated, and the biomarker or fixed fitted biomarker combination is evaluated in each bootstrap dataset.

Results

We provide three examples that illustrate the BioPET methodology. Examples 1–2 use the BioPET webtool, which analyzes a biomarker for prognostic enrichment based on its performance as summarized by the area under the receiver operating characteristic curve. Example 3 uses the BioPET R implementation to analyze a biomarker dataset.

Example 1: A moderate performance biomarker

Table 1 and Figure 1 show BioPET analysis of a biomarker with area under the receiver operating characteristic curve 0.72. This approximates the performance of a composite score using age, estimated glomerular filtration rate, and total kidney volume for prognosis of a substantial decline in renal function among patients with autosomal dominant polycystic kidney disease. The analysis assumes a 20% event rate in the non-intervention group without enrichment¹¹ and specifies 90% power to detect a 30% reduction in the event rate with $\alpha=0.025$. For illustration, the cost of screening is entered as \$1000 per patient and the average cost of a patient in the trial is entered as \$10,000.

Results (Table 1, Figure 1) show trial characteristics as a function of screening stringency. Screening stringency is represented by the percentage of patients screened out of the trial. For example, a screening threshold of 25% means that the 25% of patients at lowest risk of the endpoint are excluded from the trial. Screening threshold 0 represents no screening -- all patients are eligible.

Event rate among biomarker-positive patients gives the rate of the clinical event among “screen positive” patients without intervention. This is also the positive predictive value for the screening threshold. In Table 1, the event rate is 20% for screening threshold 0 because 20% was specified as the event rate in the population without treatment. The event rate increases with more stringent levels of screening. For example, at the 25% screening threshold the event rate is 24%; at the 75% screening threshold the event rate is 39%.

Sample size gives the result of a sample size calculation based on the event-rate among biomarker-positive patients. Table 1 shows that an unenriched trial would require 1639 patients. Using the 75% screening threshold reduces the necessary sample size to 683 because patients in the top quartile of risk have a rate of renal function decline about twice that of all patients. Sample size decreases with higher screening thresholds.

Total Screened refers to the number of patients who need to be screened to enroll the trial. Enriched trials have smaller sample sizes but also require patients to be screened to determine eligibility. For weaker biomarkers Total Screened increases with the screening threshold. For high performance biomarkers, the dramatic decrease in the trial sample size means that Total Screened actually decreases with the screening threshold (Figure 1). Importantly, Total Screened is proportional to the calendar time to enroll the trial.

Total Cost helps investigators evaluate the possible cost-savings of prognostic enrichment and whether the savings of a reduced sample size are offset by the expense of measuring the biomarker.

In Example 1, the minimum cost is attained around the 85% screening threshold, although Total Cost is similar between thresholds 75% and 90%. If the 85% threshold were used to screen patients into a trial, an adequately powered trial would require about 542 patients compared to 1639 with no screening. On average, 6.7 patients would need to be screened to identify one eligible for the trial. Using the cost estimates given above, a trial using the 85% threshold would save about half of the total patient-related costs.

Example 2: A high performance biomarker

Table 2 and Figure 1 also provide an example of a very strong marker with area under the receiver operating characteristic curve 0.92. An example of such a strong marker may be urinary albumin-creatinine ratio for prognosis of end stage renal disease or death among patients with chronic kidney disease.²¹ Aside from the area under the receiver operating characteristic curve, all inputs into the analysis are the same as Example 1. As in Example 1, with no screening the event rate is 20%. At the 25% screening threshold, the event rate is 26% for this biomarker; at the 75% screening threshold the event rate is 63%.

Table 2 and Figure 1 show that the trial sample size drops steadily as the screening threshold becomes increasingly stringent. For this high performance marker, the trial sample size decreases so dramatically that fewer total patients need to be screened to enroll an enriched trial. This implies that the calendar time to enroll an enriched trial would be shorter than for an unenriched trial.

Example 3: A moderate performance biomarker shows promise for prognostic enrichment

Using the R²² implementation of BioPET, we analyzed data on cardiac surgery patients in the Translational Research Investigating Biomarker Endpoints study^{23, 24} using the subset of the cohort (n=690) for whom 1-year mortality data were available. We consider the biomarker plasma neutrophil gelatinase-associated lipocalin, measured 0–6 hours postoperatively, to forecast 1-year mortality.

The biomarker has a modest area under the receiver operating characteristic curve of about 0.63 for death within 1 year, which occurred for 60 patients (8.7% of the sample). Table 3 and Figure 2 show the BioPET analysis results; the cost analysis assumes a \$50 cost for measuring the marker and a \$1000 cost per patient in the trial. Results indicate that the most cost-efficient trial would enroll the 30% of patients at highest risk of death (70% screening threshold). At this level of enrichment, the event rate increases from 8.7% to an estimated 14.5%. A trial enriched at this level would require about 2400 patients compared to 4000 for the unenriched trial.

Conclusions

BioPET evaluates biomarkers along different dimensions, including the trial sample size for different screening thresholds. The BioPET webtool provides easy access to a first approximation of whether a biomarker holds promise for prognostic enrichment in various clinical contexts. Results show that biomarkers that may be unimpressive in terms of area under the receiver operating characteristic curve can show promise for prognostic enrichment.

Investigators relying on published biomarker results should be cognizant of possible publication bias and other sources of optimism in those results.²⁵ We recommend a sensitivity analysis that considers varying degrees of biomarker performance. One can then examine how BioPET results change as the biomarker's performances is varied within a reasonable range. We caution investigators against assuming a particular level of biomarker performance based on a single publication.

There are practical and logistical considerations when planning any clinical trial and not all of these can be part of the BioPET quantitative analysis. For example, investigators should consider the consequences of a prognostic enrichment strategy on patient recruitment, or the effects of an extended calendar time on enrollment or staffing the trial.

The BioPET cost analysis assumes fixed costs for biomarker measurement and running a patient through a trial. The cost analysis does not reflect costs of longer or shorter enrollment periods or possible efficiencies/inefficiencies of scale. In many cases, the total cost is similar for a range of thresholds near the minimum cost threshold. We recommend that investigators consider the range of biomarker thresholds with similar Total Costs and evaluate the practical and ethical implications of choosing a higher or lower threshold in this range.

When an enriched trial is completed, questions arise about whether the results apply to patients who had been screened out of the trial.⁷ All trials have eligibility criteria, so this question is not particular to enriched trials. If an enriched trial demonstrates efficacy for the intervention, a traditional approach is to perform subsequent trials in lower risk populations.⁷ A desire for greater generalizability might lead investigators to err on the side of less stringent screening. On the other hand, ethical considerations may favor more stringent screening, so that the new treatment is tested on only the highest risk patients.

In summary, biomarkers should be evaluated in the context of how they will be used. This means going beyond measures of biomarker performance like area under the receiver operating characteristic curve that do not directly assess any clinical or public health benefit of using the biomarker. When investigators consider using a biomarker to enrich a clinical trial, they must consider multiple, sometimes-conflicting dimensions: the trial sample size, the calendar time of enrollment, the per-patient cost for the trial and for biomarker measurement, the ethics of trial eligibility criteria, and the generalizability of trial results. BioPET allows investigators to evaluate the utility of a potential prognostic enrichment biomarker using metrics that align with these considerations.

References

1. Szczech LA. The development of urinary biomarkers for kidney disease is the search for our renal troponin. *J Am Soc Nephrol.* 2009; 20:1656–1657. [PubMed: 19520749]
2. Henson DE, Srivastava S, Kramer BS. Molecular and genetic targets in early detection. *Curr Opin Oncol.* 1999; 11:419–425. [PubMed: 10505783]
3. Devarajan P. Emerging urinary biomarkers in the diagnosis of acute kidney injury. *Expert Opin Med Diagn.* 2008; 2:387–398. [PubMed: 19079800]
4. Kibe S, Adams K, Barlow G. Diagnostic and prognostic biomarkers of sepsis in critical care. *J Antimicrob Chemother.* 2011; 66(Suppl 2):ii33–ii40. [PubMed: 21398306]
5. Vasudev NS, Selby PJ, Banks RE. Renal cancer biomarkers: the promise of personalized care. *BMC Med.* 2012; 10:112. [PubMed: 23016578]
6. Yuasa T, Takahashi S, Hatake K, et al. Biomarkers to predict response to sunitinib therapy and prognosis in metastatic renal cell cancer. *Cancer Sci.* 2011; 102:1949–1957. [PubMed: 21812860]
7. Temple R. Enrichment of clinical study populations. *Clin Pharmacol Ther.* 2010; 88:774–778. [PubMed: 20944560]
8. US Food and Drug Administration. Qualification process for drug development tools. 2014. Guidance for industry and FDA staff.

9. Parikh CR, Moledina DG, Coca SG, et al. Application of new acute kidney injury biomarkers in human randomized controlled trials. *Kidney Int.* 2016; 89:1372–1379. [PubMed: 27165835]
10. Vickers AJ, Bennette C, Kibel AS, et al. Who should be included in a clinical trial of screening for bladder cancer?: a decision analysis of data from the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. *Cancer.* 2013; 119:143–149. [PubMed: 22736219]
11. Biomarker Qualification Review Team (BQRT). Biomarker qualification review for total kidney volume.
12. Fisher B, Costantino JP, Wickerham DL, et al. Tamoxifen for the prevention of breast cancer: current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study. *J Natl Cancer Inst.* 2005; 97:1652–1662. [PubMed: 16288118]
13. Kerr KF, Meisner A, Thiessen-Philbrook H, et al. Developing risk prediction models for kidney injury and assessing incremental value for novel biomarkers. *Clin J Am Soc Nephrol.* 2014; 9:1488–1496. [PubMed: 24855282]
14. Pepe MS. Receiver operating characteristic methodology. *J Am Stat Assoc.* 2000; 95:308–311.
15. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006; 26:565–574. [PubMed: 17099194]
16. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. *Lancet.* 2002; 360:7–22. [PubMed: 12114036]
17. Shepherd J, Cobbe SM, Ford I, et al. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. West of Scotland Coronary Prevention Study Group. *N Engl J Med.* 1995; 333:1301–1307. [PubMed: 7566020]
18. Downs JR, Clearfield M, Weis S, et al. Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels: results of AFCAPS/TexCAPS. Air Force/Texas Coronary Atherosclerosis Prevention Study. *JAMA.* 1998; 279:1615–1622. [PubMed: 9613910]
19. Campbell G, Ratnaparkhi MV. An application of lomax distributions in receiver operating characteristic (ROC) curve analysis. *Commun Stat Theory Methods.* 1993; 22:1681–1687.
20. Fleiss, JL. Statistical methods for rates and proportions. 2. New York, NY: Wiley, John and Sons, Incorporated; 1981.
21. Pavkov ME, Knowler WC, Hanson RL, et al. Predictive power of sequential measures of albuminuria for progression to ESRD or death in Pima Indians with type 2 diabetes. *Am J Kidney Dis.* 2008; 51:759–766. [PubMed: 18436086]
22. R development core team. R: A language and environment for statistical computing. 2013.
23. Parikh CR, Coca SG, Thiessen-Philbrook H, et al. Postoperative biomarkers predict acute kidney injury and poor outcomes after adult cardiac surgery. *J Am Soc Nephrol.* 2011; 22:1748–1757. [PubMed: 21836143]
24. Coca SG, Garg AX, Thiessen-Philbrook H, et al. Urinary biomarkers of AKI and mortality 3 years after cardiac surgery. *J Am Soc Nephrol.* 2014; 25:1063–1071. [PubMed: 24357673]
25. Kerr KF, Meisner A, Thiessen-Philbrook H, et al. RiGoR: reporting guidelines to address common sources of bias in risk model development. *Biomark Res.* 2015; 3:2. [PubMed: 25642328]

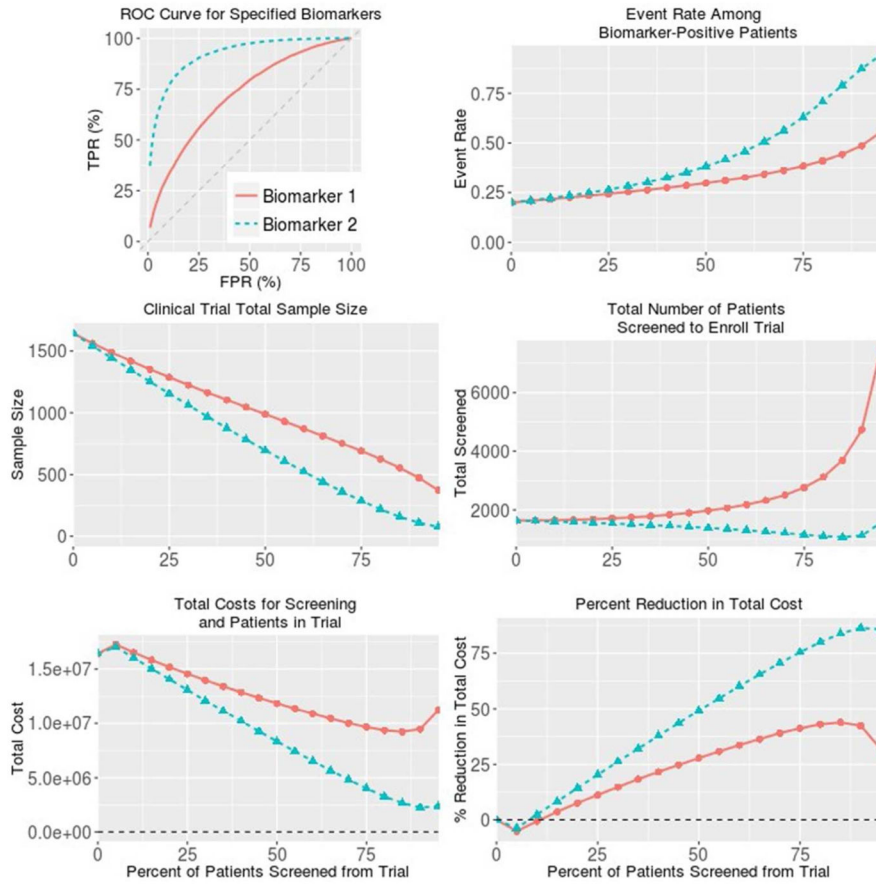


Figure 1. BioPET analysis of two biomarkers with modest (0.72; Biomarker 1) and strong (0.92; Biomarker 2) values for the area under the receiver operating characteristic curve
 The context is a clinical event occurring in 20% of patient without intervention and without biomarker screening. A clinical trial for an intervention is planned to have 90% power to detect a 30% relative reduction in the event rate using one-sided hypothesis testing and $\alpha=0.025$. For all plots except for the receiver operating characteristic curve, the horizontal axis is the percent of patients screened out of the trial.

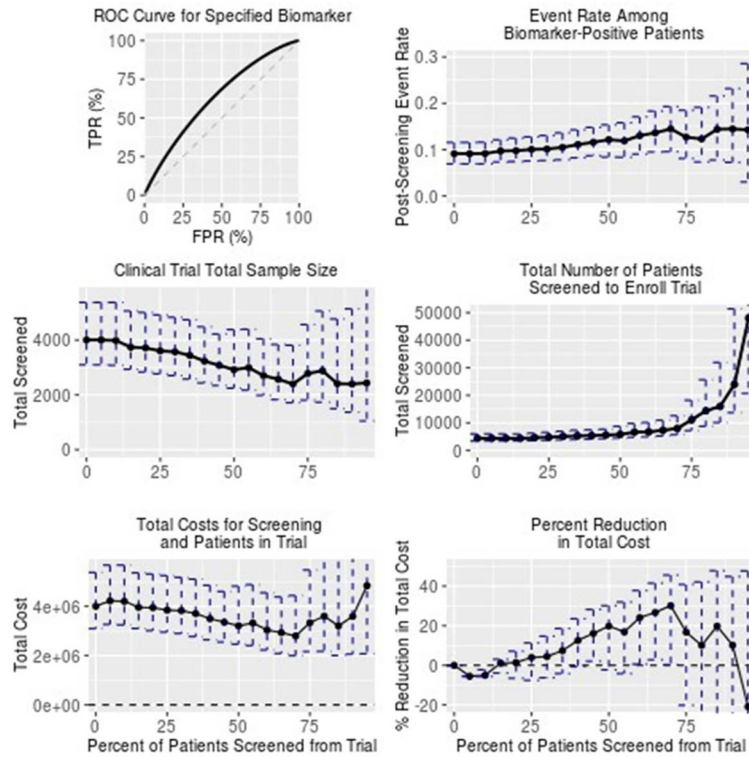


Figure 2. Evaluation of postoperative plasma neutrophil gelatinase-associated lipocalin as a prognostic enrichment biomarker for the outcome of death within one year of surgery Biomarker performance and rates of death are estimated from the Translational Research Investigating Biomarker Endpoints study data. Results here show 95% bootstrap intervals to describe uncertainty in the results. This analysis is intended for illustrative purposes rather than to report any findings of the Translational Research Investigating Biomarker Endpoints study.

Table 1
BioPET analysis of a biomarker with area under the receiver operating characteristic curve =0.72

The clinical event occurs in 20% of patients without intervention. A trial of an intervention to help prevent the clinical event is designed to have 90% power to detect a 30% relative reduction in the event rate using one-sided hypothesis testing and $\alpha=0.025$. The cost of screening is \$1000 and the per patient trial cost is \$10,000. The **Screening Threshold** is the proportion of patients who will be screened out of the trial. The **Event Rate** is the rate of the clinical event in the enriched study population not receiving the intervention. The **Sample Size** is the trial sample size calculated as a function of the event rate and statistical testing specifications. **Total Screened** is the total number of patients who would need to be screened to enroll the trial, which depends on the sample size and stringency of screening. **Total Cost** summarizes patient-related costs of different levels of enrichment, specifically the costs of biomarker-based screening and the costs of having a patient in a trial. The results here are also displayed in Figure 1.

| Screening Threshold | Event Rate | Sample Size | Total Screened | Total Cost | Percent Reduction in Total Cost |
|---------------------|------------|-------------|----------------|------------|---------------------------------|
| 0% | 0.20 | 1643 | 1643 | 16431723 | Ref |
| 5% | 0.21 | 1562 | 1645 | 17268036 | -5.1% |
| 10% | 0.22 | 1488 | 1653 | 16530257 | -0.6% |
| 15% | 0.23 | 1418 | 1669 | 15853049 | 3.5% |
| 20% | 0.23 | 1352 | 1690 | 15210458 | 7.4% |
| 25% | 0.24 | 1287 | 1716 | 14584524 | 11.2% |
| 30% | 0.25 | 1225 | 1751 | 14004331 | 14.8% |
| 35% | 0.26 | 1165 | 1792 | 13443614 | 18.2% |
| 40% | 0.27 | 1106 | 1843 | 12904180 | 21.5% |
| 45% | 0.29 | 1047 | 1903 | 12372557 | 24.7% |
| 50% | 0.30 | 989 | 1978 | 11865727 | 27.8% |
| 55% | 0.31 | 928 | 2063 | 11348024 | 30.9% |
| 60% | 0.33 | 869 | 2172 | 10862124 | 33.9% |
| 65% | 0.34 | 811 | 2316 | 10421107 | 36.6% |

| Screening Threshold | Event Rate | Sample Size | Total Screened | Total Cost | Percent Reduction in Total Cost |
|---------------------|------------|-------------|----------------|------------|---------------------------------|
| 70% | 0.36 | 751 | 2503 | 10013428 | 39.1% |
| 75% | 0.39 | 689 | 2755 | 9640789 | 41.3% |
| 80% | 0.41 | 622 | 3108 | 9323150 | 43.3% |
| 85% | 0.44 | 552 | 3681 | 9202643 | 44.0% |
| 90% | 0.49 | 476 | 4759 | 9517265 | 42.1% |
| 95% | 0.55 | 381 | 7621 | 11431165 | 30.4% |

Table 2
BioPET analysis of a biomarker with area under the receiver operating characteristic curve =0.92

The clinical event occurs in 20% of patients without intervention. A trial of an intervention to help prevent the clinical event is designed to have 90% power to detect a 30% relative reduction in the event rate using one-sided hypothesis testing and $\alpha=0.025$. The cost of screening is \$1000 and the per patient trial cost is \$10,000. The **Screening Threshold** is the proportion of patients who will be screened out of the trial. The **Event Rate** is the rate of the clinical event in the enriched study population not receiving the intervention. The **Sample Size** is the trial sample size calculated as a function of the event rate and statistical testing specifications. **Total Screened** is the total number of patients who would need to be screened to enroll the trial, which depends on the sample size and stringency of screening. **Total Cost** summarizes patient-related costs of different levels of enrichment, specifically the costs of biomarker-based screening and the costs of having a patient in a trial. The results here are also shown in Figure 1.

| Screening Threshold | Event Rate | Sample Size | Total Screened | Total Cost | Percent Reduction in Total Cost |
|---------------------|------------|-------------|----------------|------------|---------------------------------|
| 0% | 0.20 | 1640 | 1640 | 16398000 | Ref |
| 5% | 0.21 | 1541 | 1622 | 17034789 | -3.9% |
| 10% | 0.22 | 1443 | 1603 | 16034949 | 2.2% |
| 15% | 0.24 | 1346 | 1583 | 15041843 | 8.3% |
| 20% | 0.25 | 1250 | 1562 | 14059403 | 14.3% |
| 25% | 0.27 | 1155 | 1539 | 13084639 | 20.2% |
| 30% | 0.28 | 1060 | 1514 | 12109590 | 26.2% |
| 35% | 0.30 | 966 | 1487 | 11149298 | 32% |
| 40% | 0.33 | 875 | 1458 | 10204768 | 37.8% |
| 45% | 0.35 | 784 | 1426 | 9266357 | 43.5% |
| 50% | 0.38 | 695 | 1391 | 8344788 | 49.1% |
| 55% | 0.42 | 608 | 1352 | 7435845 | 54.7% |
| 60% | 0.46 | 524 | 1310 | 6548590 | 60.1% |
| 65% | 0.51 | 443 | 1264 | 5689500 | 65.3% |

| Screening Threshold | Event Rate | Sample Size | Total Screened | Total Cost | Percent Reduction in Total Cost |
|---------------------|------------|-------------|----------------|------------|---------------------------------|
| 70% | 0.56 | 364 | 1213 | 4850961 | 70.4% |
| 75% | 0.63 | 289 | 1157 | 4049439 | 75.3% |
| 80% | 0.71 | 221 | 1104 | 3313108 | 79.8% |
| 85% | 0.79 | 162 | 1077 | 2693387 | 83.6% |
| 90% | 0.87 | 113 | 1130 | 2260041 | 86.2% |
| 95% | 0.95 | 77 | 1550 | 2324983 | 85.8% |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Evaluation of postoperative plasma neutrophil gelatinase-associated lipocalin as a prognostic enrichment biomarker for the outcome of death within one year of surgery. Biomarker performance and rates of death are estimated from data from the Translational Research Investigating Biomarker Endpoints study, where the area under the receiver operating characteristic curve of the biomarker is 0.63 and the clinical event occurred in 8.7% of observations. A trial of an intervention to help prevent the clinical event is designed to have 90% power to detect a 30% relative reduction in the event rate using one-sided hypothesis testing and $\alpha=0.025$. The cost of screening is \$50 and the per patient trial cost is \$1,000. The **Screening Threshold** is the proportion of patients who will be screened out of the trial. The **Event Rate** is the rate of the clinical event in the enriched study population not receiving the intervention. The **Sample Size** is the trial sample size calculated as a function of the event rate and statistical testing specifications. **Total Screened** is the total number of patients who would need to be screened to enroll the trial, which depends on the sample size and stringency of screening. **Total Cost** summarizes patient-related costs of different levels of enrichment, specifically the costs of biomarker-based screening and the costs of having a patient in a trial. The results here are also shown in Figure 2.

Table 3

| Screening Threshold | Event Rate | Sample Size | Total Screened | Total Cost | Percent Reduction in Total Cost |
|---------------------|------------|-------------|----------------|------------|---------------------------------|
| 0% | 0.09 | 4004 | 4427 | 4003965 | 0% |
| 5% | 0.09 | 4004 | 4427 | 4225338 | -5.5% |
| 10% | 0.09 | 3983 | 4426 | 4204355 | -5% |
| 15% | 0.10 | 3739 | 4403 | 3959482 | 1.1% |
| 20% | 0.10 | 3716 | 4645 | 3948378 | 1.4% |
| 25% | 0.10 | 3605 | 4811 | 3845575 | 4% |
| 30% | 0.10 | 3571 | 5102 | 3826289 | 4.4% |
| 35% | 0.10 | 3442 | 5301 | 3707178 | 7.4% |
| 40% | 0.11 | 3231 | 5385 | 3500209 | 12.6% |
| 45% | 0.12 | 3078 | 5603 | 3357747 | 16.1% |
| 50% | 0.12 | 2919 | 5838 | 3210983 | 19.8% |
| 55% | 0.12 | 2995 | 6645 | 3327176 | 16.9% |
| 60% | 0.13 | 2702 | 6754 | 3039412 | 24.1% |

| Screening Threshold | Event Rate | Sample Size | Total Screened | Total Cost | Percent Reduction in Total Cost |
|---------------------|------------|-------------|----------------|------------|---------------------------------|
| 65% | 0.14 | 2569 | 7326 | 2935680 | 26.7% |
| 70% | 0.14 | 2397 | 7991 | 2796935 | 30.1% |
| 75% | 0.13 | 2780 | 11087 | 3334256 | 16.7% |
| 80% | 0.12 | 2881 | 14404 | 3600894 | 10.1% |
| 85% | 0.14 | 2411 | 15993 | 3210276 | 19.8% |
| 90% | 0.14 | 2397 | 23974 | 3596059 | 10.2% |
| 95% | 0.14 | 2437 | 48045 | 4839320 | -20.9% |