



HHS Public Access

Author manuscript

Ann Epidemiol. Author manuscript; available in PMC 2016 December 01.

Published in final edited form as:

Ann Epidemiol. 2015 December ; 25(12): 901–906.e4. doi:10.1016/j.annepidem.2015.08.008.

Quantifying the Value of Biomarkers for Predicting Mortality

Noreen Goldman^a and Dana A. Gleit^{b,*}

Noreen Goldman: ngoldman@princeton.edu; Dana A. Gleit: dag77@georgetown.edu

^aOffice of Population Research, Princeton University, 243 Wallace Hall, Princeton, NJ, 08544-2091, USA

^bCenter for Population and Health, Georgetown University, 37th & O Streets, NW, 309 Healy Washington, DC 20057-1197, USA

Abstract

Purpose—In light of widespread interest in the prognostic value of biomarkers, we apply three discrimination measures to evaluate the incremental value of biomarkers –beyond self-reported measures – for predicting all-cause mortality. We assess whether all three measures –AUC, NRI(>0), and IDI – lead to the same conclusions.

Methods—We use longitudinal data from a nationally representative sample of older Taiwanese ($n = 639$, aged 54+ in 2000, examined in 2000 and 2006, with mortality follow-up through 2011). We estimate age-specific mortality using a Gompertz hazard model.

Results—The broad conclusions are consistent across the three discrimination measures and support the inclusion of biomarkers, particularly inflammatory markers, in household surveys. Although the rank ordering of individual biomarkers varies across discrimination measures, the following is true for all three: interleukin-6 is the strongest predictor, the other three inflammatory markers make the top 10, and homocysteine ranks second or third.

Conclusions—The consistency of most of our findings across metrics should provide comfort to researchers using discrimination measures to evaluate the prognostic value of biomarkers. However, because the degree of consistency varies with the level of detail inherent in the research question, we recommend that researchers confirm results with multiple discrimination measures.

Keywords

Biological markers; Discrimination; Inflammation; Mortality; Prognosis; Taiwan

INTRODUCTION

Epidemiologists and clinicians have a long-standing interest in identifying biomarkers that have prognostic value for risk stratification and for predicting health events or death. Here

*Correspondence to: Dana A. Gleit, 5985 San Aleso Court, Santa Rosa, CA, 95409-3912, USA (dag77@georgetown.edu). Phone: 1 (707) 539-5592. Fax: 1 (707) 978-3213.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

we broaden this conventional inquiry by using three measures to examine the incremental value of an extensive set of biomarkers in predicting survival, above and beyond self-reports of health and socio-demographic information. Since the 1990s, an increasing number of population-based social surveys have collected biological measures alongside detailed household questionnaires that obtain information on health and disability. Unfortunately, evaluations of the usefulness of these data collection efforts are seriously lacking, a critical concern in light of the financial costs, logistic complications, respondent burden, ethical concerns and threats to privacy of the data.

Determining the incremental value of biomarkers for risk assessment is not straightforward. Statistical significance is insufficient because it is strongly influenced by sample size and fails to capture substantive importance. Epidemiologists and clinicians have relied primarily on measures of discrimination: determining how well a regression model distinguishes individuals who experience an event from those who do not. The measure used most frequently is the area under the receiver-operating-characteristic curve (AUC). Because the focus of research has generally been on the *incremental* value of biomarkers, the corresponding metric has been the change in AUC (Δ AUC) attributable to the biomarker(s) of interest.

A serious limitation of Δ AUC – its strong dependence on the strength of the baseline model – has resulted in very small improvements in the AUC when a marker has been added to a baseline model that discriminates very well (e.g., AUC > 0.80). In the case of cardiovascular risk prediction, standard as well as novel markers often have only marginal effects on Δ AUC [1, 2]. Alternative methods have been proposed to compare predictive risk models. What remains unclear is the extent to which different measures yield similar conclusions. Pencina and colleagues underscore the distinct strengths and weaknesses of various discrimination measures and argue for presentation of multiple measures when assessing the incremental predictive value of novel markers [3]. Steyerberg and colleagues demonstrate that different discrimination measures favor the inclusion of different markers [4]. Such findings are disturbing because there are no guidelines for identifying the preferred metric: these discrimination measures do not have any clinical interpretation or clinically-based cutoff values [5] and consensus for quantifying improvements in risk prediction is lacking [6]. In addition, how the measures rank various models likely depends on the particular research question.

In this paper we apply three discrimination measures to evaluate the prognostic value of a set of biomarkers collected in the Social Environment and Biomarkers of Aging Study (SEBAS) in Taiwan [7], a pioneering survey of older adults. Similar markers have been included in other recent biosocial surveys. We consider four questions, which are similar to those evaluated using only the AUC in a previous study [8]. First, do biomarkers have incremental value after adjusting for extensive self-reported information? Second, do *changes* in biomarker values provide better discrimination than a one-time measurement? Third, which cluster of biomarkers – standard cardiovascular/metabolic, inflammatory, or neuroendocrine – provides the strongest prediction? Finally, we pose a more nuanced question: which individual biomarkers are the strongest predictors? Our objective is to

assess the extent to which three discrimination measures provide consistent answers to these important issues.

MATERIAL AND METHODS

Data

Our data come from a cohort study in Taiwan, a population whose life expectancy and cause of death structure are similar to those observed in other industrialized countries including the US [9–11]. The SEBAS cohort was based on a nationally representative sample of Taiwanese aged 54 in 2000, selected randomly using a multi-stage sampling design with oversampling of older persons (71+) and urban residents [12]. In 2000, in-home interviews were completed with 1,497 respondents, 1,023 of whom completed a physical examination. Exam participants did not differ significantly from nonparticipants in ways likely to introduce serious bias [13]. Six years later, a follow-up was conducted with those who completed the 2000 exam and survived to 2006: Among 846 survivors, 757 completed the in-home interview and 639 participated in the physical examination.

The physical examination followed a similar protocol in both waves. Several weeks after the household interview, participants collected a 12-hour overnight urine sample (7pm to 7am), fasted overnight, and visited a nearby hospital the following morning for a physical examination. Union Clinical Laboratories in Taipei analyzed the blood and urine specimens; a sample of duplicate specimens was sent to Quest Diagnostics in the US for comparison. Details regarding response rates, sample attrition, exam participation, intra-lab reliability, inter-lab correlations, and compliance are provided elsewhere [12].

Survival status as of January 1, 2012 was ascertained by linkage to the death certificate file maintained by the Taiwan Department of Health and to the household registration database maintained by the Ministry of the Interior. The analysis sample was based on the longitudinal cohort that completed the exam in 2000 and 2006 ($n=639$, 104 deaths by December 31, 2011). The mean length of mortality follow-up was 5.1 years (range = 4.95–5.32). Because 89 respondents were missing data for at least one covariate, we followed standard practices of multiple imputation [14, 15] based on five imputed datasets.

Biomarkers and Control Variables

Biomarkers—We include 19 biomarkers shown by prior studies to be associated with all-cause mortality. They comprise three clusters of biologically-related markers: 1) eight standard cardiovascular/metabolic risk factors—systolic and diastolic blood pressure, high-density lipoprotein cholesterol (HDL), ratio of total to HDL cholesterol, triglycerides, glycosylated hemoglobin, body mass index, and waist circumference; 2) four inflammatory markers—interleukin-6 (IL-6), high-sensitivity C-reactive protein (CRP), soluble intercellular adhesion molecule 1 (sICAM-1), and soluble E-selectin; and 3) four neuroendocrine markers—dehydroepiandrosterone sulfate (DHEAS), cortisol, epinephrine, and norepinephrine. We also include three markers that do not represent a common biological subsystem—creatinine clearance, albumin, and homocysteine.

Self-reported health indicators—We selected six self-reported measures from 2006 that have been included in prognostic indexes for community-based populations (see Yourman et al. [16]): 1) global self-assessed health (“Regarding your current state of health, do you feel it is excellent, good, average, not so good, or poor?”); 2) an index of mobility limitations; 3) whether the respondent was ever diagnosed with diabetes; 4) history of cancer; 5) number of hospitalizations in the past 12 months; and 6) smoking status (never, former, current). The mobility index is based on self-reported difficulty performing eight physical tasks without assistance, calculated as described in Long and Pavalko [17]. With regard to self-reported diseases, we included only cancer and diabetes because these conditions are generally reported more accurately than others (see, for example, Goldman et al. [13]).

Social and demographic characteristics—We control for key demographic and social characteristics known to predict mortality. Demographic variables comprise age, sex, urban residence, and ethnicity (Mainlander vs. Taiwanese). Social measures comprise educational attainment, social integration, and perceived social support. An index of social integration, based on 10 indicators, is constructed following the strategy of Cornwell and Waite [18]. An index of perceived social support is based on four questions regarding potential instrumental and emotional support from family and friends. Additional details are provided in Table 1.

Measures of Discrimination

The most common approach for quantifying a model’s predictive power is the C-statistic or area under the receiver-operating-characteristic curve (AUC). The receiver-operating-characteristic curve is a plot of true positive rates (sensitivity) against false positive rates (1-specificity) for all possible cutoff values that discriminate between two groups (e.g., decedents vs. survivors). The AUC can be interpreted as the probability that the model predicts a higher probability of death for decedents than for survivors [19]. An AUC of 0.5 indicates that the model performs no better than chance, whereas an AUC of 1.0 represents perfect accuracy. Medical researchers have gravitated toward reclassification methods largely because the AUC is insensitive to the inclusion of a novel biomarker if the baseline model possesses good discrimination, even if the effect size is large. Reclassification methods assess the extent to which markers added to a risk model improve the classification of individuals [1, 20]. The traditional Net Reclassification Improvement (NRI) requires clinically meaningful risk strata, but we employ a newer category-free version, $NRI(>0)$, which quantifies the correct movement of model-based probabilities when additional markers are added to the model: upward for decedents and downward for survivors [21]. Unlike the AUC, the NRI depends primarily on the effect sizes of the new markers rather than the strength of the baseline model [3]. An alternative measure, the Integrated Discrimination Improvement (IDI), can be interpreted as the difference in discrimination slopes of models with and without the new markers [22], where the discrimination slope is the absolute difference in the average prediction between those who experienced the event and those who did not [23]. In contrast to the AUC and the $NRI(>0)$, the IDI incorporates information about the magnitude of change in the probabilities; it is also less dependent on

the strength of the baseline model than the AUC (but more so than the NRI(>0)) [3, 24]. See Supplementary Material for more details regarding these measures.

Analytic Strategy

Descriptive statistics (Tables 1 and 2) are weighted to account for the sampling design and for differential response rates by various covariates. Using unweighted data, we estimate a series of Gompertz hazard models, with time measured by age. Because initial tests revealed evidence of non-proportional hazards (i.e., the effect varies by age) for perceived social support, current smoker, and change (2000–06) in DHEAS, we include interactions between these variables and age.

In order to compare effect sizes across predictors, we standardize (mean=0, standard deviation=1) all continuous measures prior to model fitting. We transform biomarkers with a skewed distribution using a logarithm or power transformation (see Table 2) to better approximate normality, which substantially improves the model fit.

To address the questions presented above, we examine the improvement in discrimination – assessed by AUC, NRI(>0), and IDI – based on a comparison of models with and without the indicators being evaluated. We use the coefficients from each model to compute the predicted probability of dying by the end of follow-up for each respondent. The discrimination measures are calculated from the predicted probabilities and the observed binary outcome (death vs. survival); see Supplementary Material for details.

RESULTS

Table 3 shows the three discrimination measures for selected models; see Table S4 for measures related to relative goodness of fit (e.g., likelihood ratio test) for these same models. Although there are no established benchmarks, Pencina and colleagues suggest AUC>0.01 represents a meaningful improvement, while NRI(>0) greater than 0.6 indicates a strong contribution and NRI(>0) between 0.2 and 0.6 implies moderate improvement [3, 22]. We use these somewhat arbitrary values as benchmarks. Researchers do not provide a corresponding gauge for IDI. However, IDI values can be interpreted as the increase in average sensitivity (given fixed specificity).

Do biomarkers retain incremental prognostic value beyond self-reports?

A comparison of Models 1 and 2 suggests that biomarkers (measured in 2006) yield substantial incremental value in predicting mortality during 2006–11 beyond that of self-reported health variables: AUC= 0.04, NRI(>0)=0.74, and IDI=0.09.

Do changes in biomarkers yield better discrimination than one-time measurement?

A comparison of Models 2 and 3 reveals that the addition of biomarkers in 2000 – i.e., incorporating *change* in biomarker values – yields moderate improvement: AUC=0.02, NRI(>0)=0.56, and IDI=0.07.

Which cluster of biomarkers is the strongest predictor?

We evaluate the contributions of eight cardiovascular/metabolic markers (Model 4a), four inflammatory markers (Model 4b), and four neuroendocrine markers (Model 4c) by comparing each with Model 1. All three discrimination measures suggest that inflammatory markers yield the most predictive power.

Which individual biomarkers are the strongest predictors?

Using Model 1 as the baseline, we assess the contribution of each biomarker by adding the 2006 level and 2000–06 change for that marker. Figure 1 shows the top 10 biomarkers ranked by AUC, NRI(>0), and IDI.

For each discrimination measure, IL-6 is the strongest predictor and all four inflammatory markers make the top 10; sICAM-1 consistently has a high ranking. Although CRP is the only one of these markers used clinically, it ranks lowest of the four inflammatory markers.

The cardiovascular/metabolic markers in the top 10 generally rank near the bottom, and none ranks in the top 10 by all discrimination measures. Systolic blood pressure ranks within the top 10 on two discrimination measures; other cardiovascular/metabolic markers do so only for AUC (glycosylated hemoglobin) or IDI (HDL, body mass index, ratio total cholesterol to HDL).

DHEAS is the only neuroendocrine marker on all three lists. Epinephrine has a top 10 ranking for two of these, whereas norepinephrine and cortisol appear on the list for only one discrimination measure.

Among the three unrelated markers, homocysteine ranks second or third on each list. Creatinine clearance attains the top 10 ranking for two discrimination measures, although it fails to rank higher than 8th. Serum albumin never appears in the top 10.

When we evaluate each biomarker relative to the benchmarks, four biomarkers yield a meaningful improvement in AUC (>0.01): IL-6, DHEAS, homocysteine, and sICAM-1. One, IL-6, makes a strong contribution based on the NRI(>0); another six (homocysteine, sICAM-1, soluble E-selectin, DHEAS, systolic blood pressure, and CRP) yield a moderate improvement. These seven biomarkers also produce >1% improvement in sensitivity based on the IDI. When we examine the robustness of our findings by excluding 34 respondents with CRP > 10 mg/L (indicative of acute infection), we find most of the results unchanged.

DISCUSSION

Ascertaining whether particular biomarkers enhance prediction of health or mortality above and beyond conventional factors has been contentious. Although there are numerous statistical criteria that need to be satisfied at an early stage of analysis, most researchers ultimately focus on discrimination: does the marker improve our ability to distinguish between those who experience the event and those who do not? The merits of alternative measures of discrimination have been frequently debated, but as yet, there is no consensus about which measure is “best.” Despite a large literature on evaluating novel markers,

surprisingly few studies have assessed the consistency of findings using different discrimination measures (see Pickering and Endre [25] for an exception).

In this paper, we evaluate the robustness of conclusions about the utility of a set of biomarkers for predicting five-year all-cause mortality in a general older population based on three frequently used discrimination measures. Several broad conclusions are consistent across the AUC, NRI(>0), and IDI: (1) inclusion of biomarkers substantially enhances five-year survival prediction from a baseline model that incorporates self-reported indicators of health; (2) inclusion of information on changes in biomarkers yields a moderate improvement over one-time measurement; and (3) when considered as clusters, inflammatory markers offer stronger prediction than either cardiovascular/metabolic or neuroendocrine measures.

When we address a more specific question – which biomarkers are the strongest predictors – the findings are more nuanced. Still, all three discrimination measures underscore the utility of inflammatory markers and homocysteine levels. Standard clinical markers that reflect lipids, obesity, and blood pressure generally have relatively low prognostic power for all-cause mortality, although their predictive power would likely be higher for cardiovascular mortality.

At the same time, differences are apparent. For example, AUC and to a lesser extent NRI(>0) favor the inclusion of neuroendocrine over cardiovascular/metabolic markers (despite only half as many variables in the former category), whereas IDI favors cardiovascular/metabolic markers. We see this difference despite the fact that none of the discrimination measures penalizes for the number of parameters. The inconsistency partly reflects AUC disproportionately weighting high levels of sensitivity [22], where neuroendocrine markers outperform cardiovascular/metabolic factors. At lower levels of sensitivity, the cardiovascular/metabolic factors generally perform better than the neuroendocrine markers (as evidenced by an inspection of the ROC curves; not shown). The rank ordering of the 10 most predictive biomarkers also varies across discrimination measures.

One important limitation of this analysis is that the benchmarks we use for the discrimination measures were originally intended for testing a *single* marker. For example, if an NRI(>0) of 0.60 is indicative of a strong contribution for one marker, what magnitude should be used for 19 markers? In principle, we could impose a penalty for the number of parameters added to the model. In addition, because previous research demonstrates that AUC depends on the strength of the baseline model [3], we could permit the benchmark to vary accordingly, although there is little guidance on how to do so.

It is important to bear in mind that this study has not considered outcomes beyond all-cause mortality or the many other uses of biomarkers within population-based household surveys. The value of including biomarkers undoubtedly varies across health outcomes as well as research objectives. Moreover, the cost, complex logistics and additional burden posed by the inclusion of biomarkers in large surveys of the general population must be borne in mind. Still, at least from the point of view of predicting mortality – a well-measured

outcome highly correlated with myriad health measures – these findings provide strong support for biomarker collection within household surveys and moderate support for longitudinal collection of such markers – particularly with respect to inflammatory markers.

CONCLUSIONS

On the one hand, the consistency of most of these results across the three discrimination measures should provide some comfort to researchers. On the other hand, our findings should not signal that future evaluations based on multiple discrimination measures are superfluous. As we have shown, the degree of consistency across metrics varies with the level of detail inherent in the question. Researchers would be wise to confirm findings with multiple metrics, for example, considering AUC >0.01 as a necessary but not sufficient benchmark, and using the NRI(>0) and the IDI as robustness checks. If all three discrimination measures yield similar conclusions, the utility of the biomarker is on much firmer ground than if the metrics produce disparate results.

Upon discovering that most published research findings are inaccurate and that many of the associations reported in highly cited biomarker studies are exaggerated, Ioannidis and colleagues argue that “the standards for claiming success should be higher” [26, 27]. This analysis provides one small step in the right direction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

ACKNOWLEDGMENTS AND FUNDING

This work was supported by the National Institute on Aging (grant numbers R01AG16790, R01AG16661) and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (grant number P2CHD047879). Funding for the parent study (TLISA) came from the Taiwan Department of Health, the Taiwan National Health Research Institute [grant number DD01-86IX-GR601S] and the Taiwan Provincial Government. SEBAS was funded by the Demography and Epidemiology Unit of the Behavioral and Social Research Program of the National Institute on Aging [grant numbers R01 AG16790, R01 AG16661]. The Bureau of Health Promotion (BHP, Department of Health, Taiwan) provided additional financial support for SEBAS 2000. The sponsors had no involvement in the study design, data collection, analysis, interpretation of the data, writing of the report, or the decision to submit the article for publication.

We acknowledge the hard work and dedication of the staff at the Center for Population and Health Survey Research (BHP), who were instrumental in the design and implementation of the SEBAS and supervised all aspects of the fieldwork and data processing. We are also grateful to: Dr. Maxine Weinstein for her helpful comments and suggestions regarding this manuscript; and Dr. Germán Rodríguez for his statistical advice and programming assistance.

List of Abbreviations

AUC	area under the receiver-operating-characteristic curve
BMI	Body Mass Index
CrCl	Creatinine Clearance

CRP	C-reactive protein
DHEAS	dehydroepiandrosterone sulfate
HbA1c	Glycosylated Hemoglobin
Hcy	Homocysteine
HDL	high-density lipoprotein cholesterol
IDI	Integrated Discrimination Improvement
IL-6	interleukin-6
NRI(>0)	Continuous Net Reclassification Improvement
SBP	Systolic Blood Pressure
sE-selectin	soluble E-selectin
sICAM-1	soluble intercellular adhesion molecule 1
TC/HDL	ratio of total cholesterol to HDL

References

1. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007; 115(7):928–935.10.1161/CIRCULATIONAHA.106.672402 [PubMed: 17309939]
2. Cook NR, Buring JE, Ridker PM. The effect of including C-reactive protein in cardiovascular risk prediction models for women. *Ann Intern Med*. 2006; 145(1):21–29. [PubMed: 16818925]
3. Pencina MJ, D'Agostino RB, Pencina KM, Janssens ACW, Greenland P. Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol*. 2012; 176(6):473–481.10.1093/aje/kws207 [PubMed: 22875755]
4. Steyerberg EW, Pencina MJ, Lingsma HF, Kattan MW, Vickers AJ, Van Calster B. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. *Eur J Clin Invest*. 2012; 42(2):216–228.10.1111/j.1365-2362.2011.02562.x [PubMed: 21726217]
5. Kerr KF, Bansal A, Pepe MS. Further insight into the incremental value of new markers: the interpretation of performance measures and the importance of clinical context. *Am J Epidemiol*. 2012; 176(6):482–487.10.1093/aje/kws210 [PubMed: 22875756]
6. Pepe MS, Janes H. Commentary: Reporting standards are needed for evaluations of risk reclassification. *Int J Epidemiol*. 2011; 40(4):1106–1108.10.1093/ije/dyr083 [PubMed: 21571811]
7. Cornman JC, Gleib DA, Goldman N, et al. Cohort profile: the Social Environment and Biomarkers of Aging Study (SEBAS) in Taiwan. *International Journal of Epidemiology*. published online ahead of print September 8, 2014. 10.1093/ije/dyu179
8. Gleib DA, Goldman N, Rodríguez G, Weinstein M. Beyond self-reports: changes in biomarkers as predictors of mortality. *Population and Development Review*. 2014; 40(2):331–360. [PubMed: 25089065]
9. University of California, Berkeley (USA); Max Planck Institute for Demographic Research (Germany). [Accessed May 13, 2015] Human Mortality Database. www.mortality.org. Updated 2015
10. Department of Health, Executive Yuan, R.O.C. (Taiwan). [Accessed February 27, 2013] Deaths in Taiwan. 2011. <http://www.doh.gov.tw/EN2006/DisplayFile.aspx?url=http%3a%2f%2fwww.doh.gov.tw%2fufile%2fdoc%2fDeaths+in+Taiwan%2c+2011-20121017.docx&name=Deaths+in+Taiwan%2c+2011-20121017.docx>. Updated 2012

11. Hoyert, DL.; Xu, J. Deaths: preliminary data for 2011. Vol. 61. Hyattsville, MD: National Center for Health Statistics; 2012.
12. Chang, M.; Lin, H.; Chuang, Y., et al. Social Environment and Biomarkers of Aging Study (SEBAS) in Taiwan, 2000 and 2006: main documentation for SEBAS longitudinal public use data (released 2012). Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]; 2012. No. ICPSR03792-v5
13. Goldman N, Lin I, Weinstein M, Lin Y. Evaluating the quality of self-reports of hypertension and diabetes. *J Clin Epidemiol.* 2003; 56(2):148–154.10.1016/S0895-4356(02)00580-2 [PubMed: 12654409]
14. Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res.* 1999; 8(1):3–15. [PubMed: 10347857]
15. Rubin DB. Multiple imputation after 18+ years (with discussion). *Journal of the American Statistical Association.* 1996; 91:473–489.
16. Yourman LC, Lee SJ, Schonberg MA, Widera EW, Smith AK. Prognostic indices for older adults: a systematic review. *JAMA.* 2012; 307(2):182–192.10.1001/jama.2011.1966 [PubMed: 22235089]
17. Long JS, Pavalko E. Comparing alternative measures of functional limitation. *Med Care.* 2004; 42(1):19–27.10.1097/01.mlr.0000102293.37107.c5 [PubMed: 14713736]
18. Cornwell EY, Waite LJ. Measuring social isolation among older adults using multiple indicators from the NSHAP study. *J Gerontol B Psychol Sci Soc Sci.* 2009; 64 (Suppl 1):i38–46.10.1093/geronb/gbp037 [PubMed: 19508982]
19. Pencina MJ, D’Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med.* 2004; 23(13):2109–2123.10.1002/sim.1802 [PubMed: 15211606]
20. Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. *Ann Intern Med.* 2009; 150(11):795–802. [PubMed: 19487714]
21. Pencina MJ, D’Agostino RBS, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med.* 2011; 30(1):11–21.10.1002/sim.4085 [PubMed: 21204120]
22. Pencina MJ, D’Agostino RB Sr, D’Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med.* 2008; 27(2):157–72. discussion 207–12. 10.1002/sim.2929 [PubMed: 17569110]
23. Yates JF. External correspondence: decomposition of the mean probability score. *Organizational Behavior and Human Performance.* 1982; 30:132–156.
24. Pencina MJ, D’Agostino RB, Demler OV, Janssens AC, Greenland P. Pencina et al. respond to “The incremental value of new markers” and “Clinically relevant measures? A note of caution”. *Am J Epidemiol.* 2012; 176(6):492–494.10.1093/aje/kws206 [PubMed: 22875757]
25. Pickering JW, Endre ZH. New metrics for assessing diagnostic potential of candidate biomarkers. *Clin J Am Soc Nephrol.* 2012; 7(8):1355–1364.10.2215/CJN.09590911 [PubMed: 22679181]
26. Ioannidis JP, Panagiotou OA. Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. *JAMA.* 2011; 305(21):2200–2210.10.1001/jama.2011.713 [PubMed: 21632484]
27. Ioannidis JP. Why most published research findings are false. *PLoS Med.* 2005; 2(8):E124.10.1371/journal.pmed.0020124 [PubMed: 16060722]

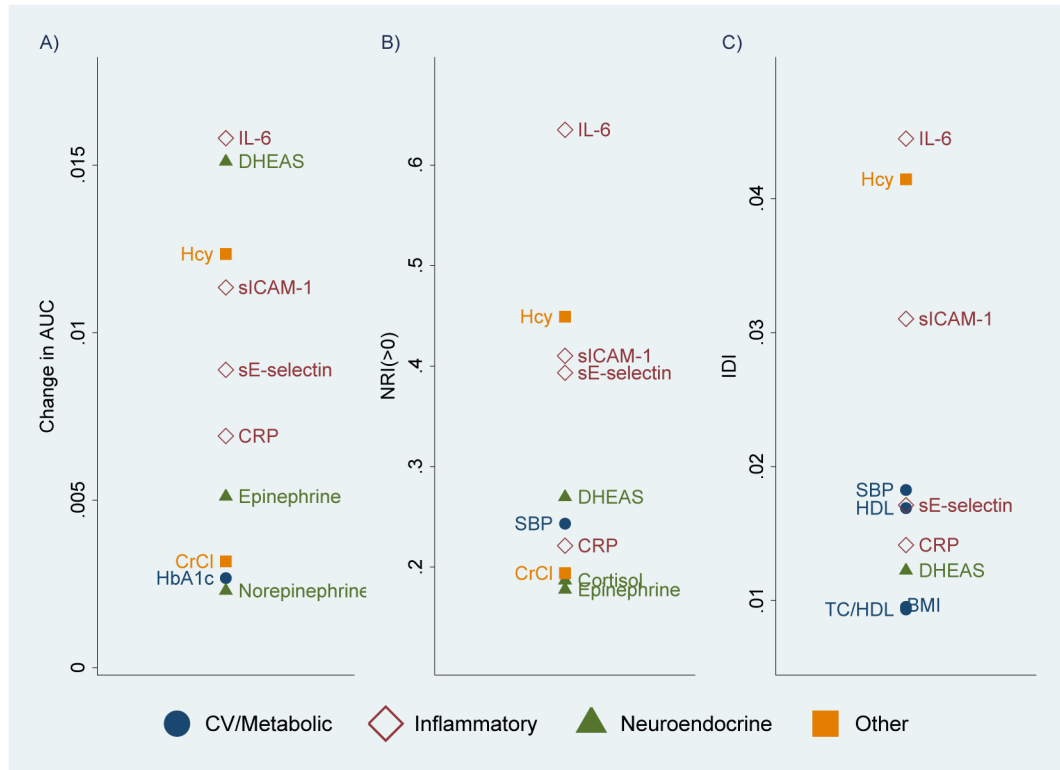


Figure 1. Top 10 Biomarkers Ranked by: A) Change in AUC; B) NRI(>0); and C) IDI, Taiwan, 2000–2011, SEBAS

Measures of discrimination were calculated based on comparisons with Model 1 (Table 3), which included social and demographic characteristics and self-reported indicators of health status. For each of the 19 biomarkers, we estimated a separate model that added the 2006 level and 2000–06 change in the specified biomarker to Model 1 to predict mortality for 2006-2011.

Table 1

Descriptive Statistics for Social and Demographic Characteristics, Self-reported Indicators of Health Status, and Survival Status, Weighted Analyses, Taiwan, 2006–2011, SEBAS

	Analysis sample (N=639)
<u>Social and demographic characteristics</u>	
Age at the 2006 exam (60–97), mean (SD)	72.0 (7.4)
Female, %	44.4
Mainlander, %	12.9
Urban resident, %	42.3
Years of completed education (0–17), mean (SD)	5.3 (4.5)
Social integration (–1.5 to 1.6), mean (SD) ^a	0.1 (0.5)
Perceived availability of social support (0.5–4.0), mean (SD) ^b	3.1 (0.7)
<u>Self-reported health indicators (2006)</u>	
Self-assessed health status (1–5, 5=excellent), mean (SD)	3.0 (1.0)
Index of mobility limitations (–0.7 to 3.2), mean (SD) ^c	0.7 (1.3)
History of diabetes, %	19.9
History of cancer, %	4.8
Number of hospitalizations in the past 12 months (0–11), mean (SD)	0.3 (0.8)
Smoking status	
Never, %	59.1
Former, %	22.2
Current, %	18.7
Died between the 2006 exam and December 31, 2011, %	16.2

^aThis index was created by standardizing each of 10 indicators (network size, network range, married/partnered, household size, does not live alone, number of friends, religious attendance, socializing with others, volunteer work, participation in social organizations) from the 2003 Taiwan Longitudinal Study of Aging (TLISA) and then calculating the mean across valid items if at least eight items were valid ($\alpha=0.72$). See Table S3 of Supplementary Material for more details.

^bEach of the following indicators was coded 0–4: family/friends willing to listen; family/friends make you feel cared for; satisfaction with emotional support received from family; can count on family to take care of you when you are ill. We calculated the mean across valid items if at least 3 items were valid ($\alpha=0.84$).

^cEach of eight tasks was coded on a four-point scale (0=no difficulty, 1=some difficulty, 2=great difficulty, 3=unable): stand for 15 minutes, squat, raise both hands overhead, grasp or turn objects with his or her fingers, lift or carry an object weighing 11–12kg, walk 200–300m, run 20–30m, and climb two or three flights of stairs. Based on the recommendations of Long and Pavalco [17], we summed the eight items (potential range 0–24), added a constant (0.5), and took the logarithm of the result to denote relative effects.

Summary Statistics for Individual Biomarkers and Changes in Biomarkers, Weighted Analyses, Taiwan, 2000–2006, SEBAS (N=639)

Table 2

		Mean (SD) for the Transformed Markers:	
	Units	Transformation	Level in 2006
			Change (2006 – 2000)
Systolic blood pressure (SBP)	mmHg	log	4.90 (0.15)
Diastolic blood pressure (DBP)	mmHg	log	4.28 (0.15)
High-density lipoprotein cholesterol (HDL)	mg/dL	log	3.84 (0.28)
Ratio of total to HDL cholesterol (TC/HDL)	ratio	log	1.43 (0.27)
Triglycerides	mg/dL	log	4.57 (0.51)
Glycosylated hemoglobin (HbA1c)	%	$-1/(\text{HbA1c})^2$	-0.03 (0.01)
Body Mass Index (BMI)	$\frac{\text{weight (kg)}}{(\text{height (m)})^2}$	log	3.20 (0.15)
Waist circumference	cm	none	84.95 (9.91)
Interleukin-6 (IL-6)	pg/mL	log	1.06 (0.80)
C-reactive protein (CRP)	mg/L	log	-2.03 (1.13)
Soluble intercellular adhesion molecule 1 (sICAM-1)	ng/mL	square root	16.50 (2.92)
Soluble E-selectin (sE-selectin)	ng/mL	log	3.57 (0.61)
Dehydroepiandrosterone sulfate (DHEAS)	µg/dL	square root	8.84 (3.22)
Cortisol	µg/g	log	2.68 (0.86)
Epinephrine	µg/g	log	1.23 (0.58)
Norepinephrine	µg/g	log	3.17 (0.53)
Creatinine Clearance (CrCl)	ml/min	none	59.65 (20.13)
Albumin	g/dL	cubed	83.65 (17.26)
Homocysteine (Hey)	µmol/L	log	2.47 (0.38)
			-0.21 (0.31)

Table 3

Measures of Discrimination for Models Predicting Mortality as a Function of Social and Demographic Characteristics, Self-reported Indicators of Health Status, and Biomarkers, Taiwan, 2006–2011, SEBAS (N=639)

Model	Description	AUC	AUC ^c	NRI(>0) ^c	IDI ^c
1	Baseline: Self-reported indicators only ^a	0.816 (0.771–0.862)			
2	Model 1 + 19 Individual biomarkers (2006) ^b	0.857 (0.817–0.897)	0.041 (0.010–0.071)	0.74 (0.51–0.97)	0.09 (0.05–0.12)
3	Model 2 + Changes in 19 Individual biomarkers (2006 and changes 2000–06) ^b	0.877 (0.839–0.914)	0.020 (–0.004 to 0.043)	0.56 (0.35–0.78)	0.07 (0.04–0.10)
4a	Model 1 + 8 Cardiovascular/metabolic markers (2006 and changes 2000–06)	0.818 (0.770–0.866)	0.001 (–0.022 to 0.025)	0.37 (0.16–0.58)	0.05 (0.03–0.08)
4b	Model 1 + 4 Inflammatory markers (2006 and changes 2000–06)	0.839 (0.795–0.882)	0.023 (–0.004 to 0.049)	0.59 (0.38–0.80)	0.07 (0.03–0.10)
4c	Model 1 + 4 Neuroendocrine markers (2006 and changes 2000–06)	0.836 (0.795–0.877)	0.019 (–0.006 to 0.045)	0.46 (0.23–0.69)	0.03 (0.01–0.05)

Abbreviations: AUC, area under the receiver-operating-characteristic curve; AUC, change in AUC; IDI, Integrated Discrimination Improvement; NRI(>0), Continuous Net Reclassification Improvement.

Note: The 95% confidence intervals are shown in parentheses next to each estimate.

^aBaseline model adjusts for: age (time-scale), sex, Mainlander, urban, education, social integration, perceived availability of support, smoking status, self-assessed health status, index of mobility limitations, history of diabetes, history of cancer, and number of hospitalizations in the past 12 months. Education was obtained when the respondent first entered the parent survey, TLISA (1989 for the oldest cohort, 1996 for the younger cohort); social integration and perceived availability of support come from the 2003 TLISA; the remainder of these control variables come from the 2006 SEBAS.

^bThe 19 biomarkers include cardiovascular/metabolic (SBP, DBP, ratio TC/HDL, HDL, triglycerides, HbA1c, BMI, waist), inflammatory (IL-6, CRP, sICAM-1, sE-selectin), and neuroendocrine markers (DHEAS, cortisol, epinephrine, norepinephrine) along with a few other markers that are unrelated biologically (creatinine clearance, serum albumin, homocysteine).

^cMeasures of the improvement in discrimination (AUC, NRI(>0), IDI) are based on comparisons with the model indicated. Note: Values for NRI(>0) and IDI are based on the average across five multiply imputed datasets.