Supplemental Material to:

Comprehensive Peroxidase-Based Hematologic Profiling for the Prediction of One-year Myocardial Infarction and Death

Marie-Luise Brennan, Ph.D.[1,2,3,#], Anupama Reddy, M.S. [2,4], W. H . Wilson Tang, M.D.[2,3,6], Yuping Wu, Ph.D. [5], Danielle M. Brennan, M.S.[6], Amy Hsu, M.S.[6], Shirley A. Mann, B.S.[1,2], Peter L. Hammer, Ph.D.[4,†], and Stanley L. Hazen, M.D., Ph.D.[1,2,3,6]

[1] Department of Cell Biology, Cleveland Clinic, Cleveland, OH
[2] Center for Cardiovascular Diagnostics and Prevention, Cleveland Clinic, Cleveland, OH
[3] Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, OH
[4] Rutgers Center for Operations Research, Rutgers University, Piscataway, NJ
[5] Department of Mathematics, Cleveland State University, Cleveland, OH
[6] Department of Cardiovascular Medicine, Cleveland Clinic, Cleveland, OH
[#] Current Address: Department of Medicine, Stanford University, Palo Alto, CA
[†] Deceased

I. General Methods and Clinical Definitions

II. Hematology Analyses and Extraction of Data Using Microsoft Excel Macro

III. Example of Peroxidase-based Flow Cytometry Cytogram

IV. Definitions of Hematology Variables

V. Example Calculation of the PEROX Risk Score

VI. PEROX Model Validation

VII. Univariate Cox Proportional Hazard Analysis for Prediction of One-Year Outcomes Using Peroxidase-based Hematology Parameters Included in PEROX Model

VIII. References

I. General Methods and Clinical Definitions

Hematology analyses were performed using an Advia 120 hematology analyzer (Siemens, New York, New York), which uses in situ peroxidase cytochemical staining to generate a CBC and differential based on flow cytometry analysis of whole anticoagulated blood. Additional white blood cell, red blood cell, and platelet related parameters derived from both cytograms and absorbance data were extracted from DAT files used in generating the CBC and differential. All hematology parameters selected for potential use in the PEROX risk score demonstrated reproducible results upon replicate (>10 times) analysis (i.e. those with a standard deviation from mean greater than 30% were excluded from inclusion in the derivation of the PEROX risk score). A blinded reviewer using established screening criteria sequentially assessed all cytograms prior to accepting specimen data. The reproducibility of the PEROX risk score was assessed by examining multiple replicate samples from multiple subjects both within and between days, revealing intra-day and inter-day coefficients of variance of 5 ± 0.4% (mean ± S.D.) and 10 ± 2%, respectively.

The mathematical method logical analysis of data[1-3] was used to identify binary variable pairs that form reproducible positive and negative predictive patterns, and to build a model predictive of risk for death or MI at one-year. Variables were included based on clinical significance, perceived potential informativeness, reproducibility (for hematology parameters) as monitored in inter-day and intra-day replicates, as well as non-redundancy, as assessed by cluster analysis performed within leukocyte, erythrocyte, and platelet subgroups. Definitions for these variables are listed below.

Criteria for the development of the PEROX risk score model included three equal proportions for each hematology parameter variable, two variables per pattern, and a minimal prevalence of 10% of the events for high-risk and 10% of non-events for low-risk patterns. Patterns were generated using logical analysis of data software (http://pit.kamick.free.fr/lemaire/LAD/), and tuned for both homogeneity and prevalence to obtain best accuracy on cross validation experiments. The weight for each positive pattern was [+1/number of high-risk patterns], while for each negative pattern was [-1/number of negative patterns]. The overall risk score a patient was assigned is calculated by the sum of positive and negative pattern weights. A maximum score of +1 would be calculated in a patient with only positive patterns whereas a maximum score of -1 would be present in a patient with only negative patterns. The original score range was adjusted from ±1 to a range of 0 to 100 by assuming 50 (rather than 0) as midpoint of equal variance. The PEROX risk score was calculated: 50 x [(1/23 possible high-risk patterns) x (# actual high-risk patterns) – (1/24 possible low-risk patterns) x (# low-risk patterns)] + 50. An example calculation is provided at the end of this Supplement.

Clinical definitions for Table 1 were defined as follows. Hypertension was defined as systolic blood pressure ≥140 mmHg, diastolic blood pressure ≥90 mmHg or taking calcium channel blocker or diuretic medications. Current smoking was defined as any smoking within the past month. History of cardiovascular disease was defined as history of cardiovascular disease, coronary artery bypass graft surgery, percutaneous coronary intervention, myocardial infarction, stroke, transient ischemic attack or sudden cardiac death. Estimated creatinine clearance was calculated using Cockcroft-Gault formula. Myocardial infarction was defined by positive cardiac enzymes, or ST changes present on electrocardiogram. Death was defined by Social Security Death Index query.

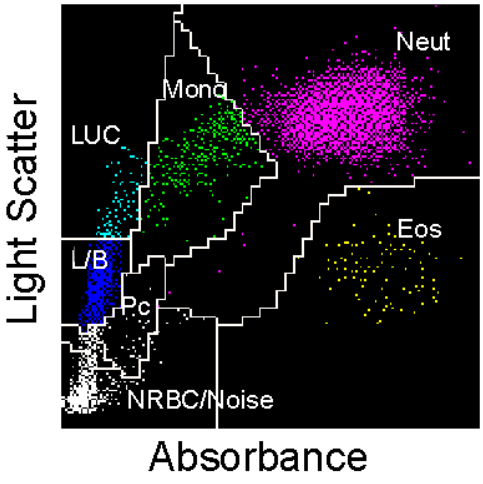
II. Hematology Analysis and Extraction of Data Using Microsoft Excel Macro

Hematology analyses were performed using an Advia 120 hematology analyzer (Siemens, New York, New York). This hematology analyzer functions as a flow cytometer, using in situ peroxidase cytochemical staining to generate a CBC and differential based on flow cytometry analysis of whole anticoagulated blood. An example of a leukocyte cytogram and a table listing all hematology analyzer elements recovered for analysis are shown below. All hematology data utilized was generated automatically by the analyzer during routine performance of a CBC and differential without any additional sample preparation or processing steps. However, additional steps must be taken to ensure the data is saved and extracted appropriately. Information on how to save and extract data is included here. Also, note that these procedures are obtainable from the instrument technical manual as part of the standard operating procedure for the machine. To improve reproducibility of hematology parameters, increased frequency of the calibrator (Cal-Chex H produced by Streck, Omaha, Nebraska) for the hematology analyzer was used (twice weekly and with reagent changes).

Data is saved by going to “Data options” tab on the Advia 120 main menu and selecting the “Data export box” (this automatically stores the hematology data in DAT files). In addition, unselect “unit set” and “unit label”. This allows for data to be collected out to additional significant digits. Data can be extracted by opening the DAT files and cutting and pasting into Microsoft Excel. Alternatively, one can use an Excel macro, a copy of which is provided as Supplemental Material available on-line (file name: ADVIA120_ExcelMacroRevF_Oct2009). To utilize the macro, the user should create two folders on the computer desktop. One should be named “export data” and the user should copy the DAT file that needs to be extracted into this folder. The other folder should be named “output data”. The user should open the macro and put the location of the export data and output data in the boxes “Export data” and “Output data”. For example if these folders are on the desktop, one would type in “c: my computer/my desktop/export data” in the “Export data” field. The user should then select “Extract data” and when prompted select the desired DAT file to be extracted. Data will then automatically be extracted with the output present as an excel file in the “Output data” folder.

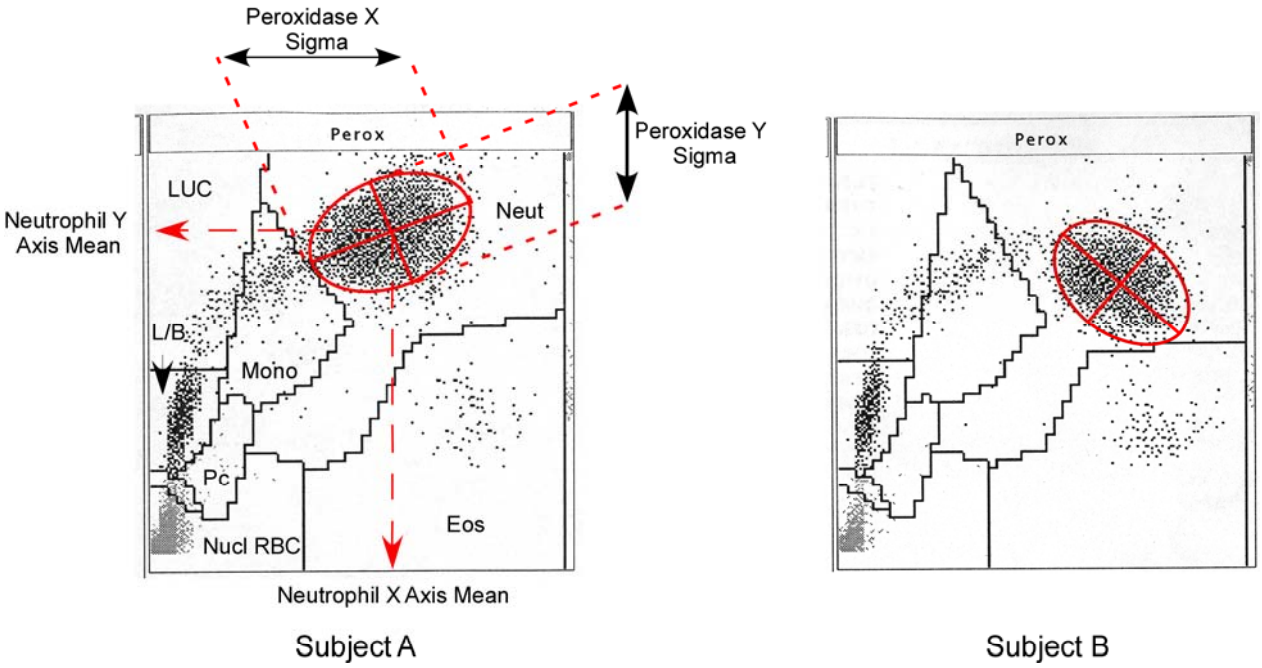
III. Sample of Peroxidase-based Flow Cytometry Cytogram

Shown below is a sample of a peroxidase-based flow-cytometry cytogram from the Advia 120 (Siemens). Light scatter measures are on Y axis (surrogate of cellular size) and absorbance measurements are on X axis (surrogate of peroxidase activity). To generate a cell count and differential, populations within pre-specified gates (shown below) are counted.



Supplemental Figure 1: Example of a Cytogram (~50,000 cells) as it appears on the analyzer screen.

Cell types are distinguished based on differences in peroxidase staining and associated absorbance and scatter measurements. Clusters are in different colors and abbreviations are included to help in distinguishing cell types. Abbreviations: Neutrophils (Neut), Monocytes (Mono), Large unstained cells (LUC), Eosinophils (Eos), Lymphocytes and basophils (L/B), Platelet clumps (Pc) and Nucleated RBCs and Noise (NRBC/Noise).



Supplemental Figure 2: Examples of Peroxidase Cytograms from 2 Subjects.

Shown are two examples of cytograms from different subjects. Some of the hematology variables related to the neutrophil main cluster are shown. Subject A has low PEROX risk score. Subject B has a high PEROX risk score. While visual inspection of the cytograms reveals clear differences, the ultimate assignment into “low” (e.g. bottom tertile) vs. “high” (top tertile) risk categories is not possible by visual inspection, since the final PEROX risk score is dependent upon the weighted presence of multiple binary pairs of low and high risk patterns derived from clinical data, laboratory data and hematological parameters from erythrocyte, leukocyte and platelet lineages. In general, cellular clusters (and subclusters) can be defined mathematically by an ellipse, with major and minor axes, distribution widths along major and minor axes, location and angles relative to the X and Y axes, etc. In addition, positional relationships between various (sub)cellular clusters can also be quantified. In this manner, multiple specific quantifiable parameters derived from the leukocyte lineage are reproducibly defined in a given peroxidase (leukocyte) cytogram. Similar phenotypic characterization of erythrocyte (predominantly determined spectrophotometrically), and platelet (cytographic analysis) lineages are also routinely collected as part of a CBC and differential. The availability of this rich array of phenotypic data as part of a routine automated CBC and differential, combined with the fact that erythrocyte, leukocyte (peroxidase) and platelet related processes are mechanistically linked to atherothrombotic disease, was the stimulus for our hypothesis that cardiovascular risk information was available within a comprehensive hematology analysis.

The final PEROX score calculation uses only a subset of hematology analyzer elements that are generated during the course of a CBC and differential, in combination with clinical and laboratory data that would routinely be available at patient encounter in an out patient setting. The table below shows only those hematology elements that are used during calculation of the PEROX risk score. Also shown are the definition of the hematology elements, and the abbreviations used within the instrument DAT files.

IV. Supplemental Table 1. Definitions of Hematology Variables

	Abbrs.	Definition
White Blood Cell Related		
White blood cell count	WBC	White blood cell count using perox methodology
Neutrophil count	#NEUT	Neutrophil cell count from neutrophil region of perox cytogram
Lymphocyte count	#LYMPH	Lymphocyte cell count from lymphocyte region of perox cytogram
Monocyte count	#MONO	Monocyte cell count from monocyte region of perox cytogram
Eosinophil count	#EOS	Eosinophil cell count from eosinophil region of perox cytogram
Basophil count	#BASO	Basophil cell count from baso region of baso cytogram
Number of peroxidase saturated cells	# PERO SAT	Number of cells in last 3 channels of perox cytogram
Neutrophil cluster mean X	NEUTX	Mean channel value of neutrophil cluster on X-axis
Neutrophil cluster mean Y	NEUTY	Mean channel value of neutrophil cluster on Y-axis
Ky	KY	Measure of fit; i.e. how well neutrophils and lymphocytes fit predicted clusters
Peroxidase X sigma	PXXSIG	Distribution width of neutrophil cell cluster; Two standard deviations from neutrophil X mean value
Peroxidase Y mean	PXY	Mean position of neutrophil cluster on Y axis; alternative measure
Peroxidase Y sigma	PXYSIG	Distribution width of neutrophil cell cluster; Two standard deviations from neutrophil Y mean value
Lobularity index	LI	Measure of white blood cell maturity; ratio of mode channels of polymorphonuclear cells per mononuclear cells
Lymphocyte/large unstained cell threshold	LUC	Highest scatter value of lymphocytes from noise/lymphocyte valley
Perox d/D	PXDD	Measure of quality of distance between lymphocyte and noise clusters
Blasts	%BLASTS	Percent of cells in blast region of basophil cytogram
Polymorphonuclear ratio		Ratio of neutrophils per eosinophils in basophil cytogram
Polymorphonuclear cluster x axis mode	PMNX	Mode of neutrophil cluster from basophil cytogram
Mononuclear central x channel	MNX	Central X channel values from basophil cytogram
Mononuclear central y channel		Central Y channel value from basophil cytogram
Mononuclear polymorphonuclear valley	MNPMN	Distance between mononuclear and polymorphonuclear clusters in basophil cytogram
Red Blood Cell Related		
RBC count	RBC	RBC counted in RBC/platelet cytogram
Hematocrit	HCT	Percent of blood consisting of RBCs; (RBC*MCV)/10
Mean corpuscular volume	MCV	Mean channel of RBC volume histogram
Mean corpuscular hemoglobin	MCH	Mean hemoglobin; calculated as hemoglobin per RBC count
Mean corpuscular hemoglobin concentration	MCHC	Mean hemoglobin concentration; Hemoglobin*1000/RBC*MCV
RBC hemoglobin concentration mean	CHCM	Mean channel of RBC hemoglobin concentration channel
RBC distribution width	RDW	Distribution width of RBC volumes; RBC volume standard deviation/MCV *100
Hemoglobin distribution width	HDW	Distribution width of RBC hemoglobin concentration; Standard deviation of hemoglobin concentration histogram
Hemoglobin content distribution width	HCDW	Standard deviation of hemoglobin content histogram
Normochromic/Normocytic RBC count		RBCs normochromic (hemoglobin concentration between 28 to 41 g/dL) and normocytic (size between 20 to 120 fL)
Macrocytic RBC count	#MACRO	RBCs with volume greater than 120 fL
Hypochromic RBC count	#HYPO	RBCs with hemoglobin concentrations less than 28 g/dL

	Abbrs.	Definition
Platelet Related		
Plateletcrit	PCT	Percent of blood consisting of platelets; MPV*PLT
Mean platelet volume	MPC	Mean platelet volume
Platelet count	PLT	Platelet count
Mean platelet component concentration	MPC	Mean of platelet component concentration
Platelet concentration distribution width	PCDW	Distribution width of platelet component concentration; two standard deviations for platelet component concentration
Large platelets	#L-PLT	Percent of platelets that are between 20 to 30 fL
Platelet clumps	PLT CLU	Percent of platelet clumps in platelet cytogram

V. Example Calculation of the PEROX Risk Score

A 62 year old stable, non-smoking, non-diabetic female with history of hypertension but no history of cardiovascular disease was seen. A CBC with differential was run. Results from a recent basic metabolic panel and fasting lipid profile are available. Blood pressure and body mass index were measured. Pertinent clinical and laboratory values are shown below.

Supplemental Table 2

Clinical and Laboratory Data	Abbr.	Value
Traditional Risk Factors		
Age (years)[†]	AGE	62
Male	MALE	No
History of Hypertension	HTN	Yes
Current smoker	SMOKE	No
Diabetes mellitus	DM	No
History cardiovascular disease	CAD	No
Laboratory Data		
Fasting blood glucose (mg/dl)	GLUC	95.2
Creatinine (mg/dl)	CREAT	0.83
Potassium (mmol/l)	K	4.0
C-reactive protein (mg/dl)	CRP	1.38
High Density Lipoprotein cholesterol (mg/dl)	HDL	44
Triglycerides (mg/dl)	TGS	161
Clinical Characteristics		
Systolic blood pressure (mm Hg)	SBP	125
Body mass index (kg/m^2)	BMI	29.0
Hematology Analyzer Data	**Abbr.**	**Value**
White Blood Cell Related		
White blood cell count (x10^3/μl)	WBC	7.34
Neutrophil count (x10^3/μl)	#NEUT	4.53
Lymphocyte count (x10^3/μl)	#LYMPH	2.10
Monocyte count (x10^3/μl)	#MONO	0.37
Eosinophil count (x10^3/μl)	#EOS	0.13
Basophil count (x10^3/μl)	#BASO	0.02
Number of peroxidase saturated cells (x10^3/μl)	#PEROXSAT	0.00
Neutrophil cluster mean x	NEUTX	64.4
Neutrophil cluster mean y	NEUTY	74.8
Ky	KY	100
Peroxidase x sigma	PXXSIG	0.00
Peroxidase y mean	PXY	19.06
Peroxidase y sigma	PXYSIG	6.55
Lobularity index	LI	0.40
Lymphocyte/large unstained cell threshold	LUC	50
Perox d/D	PXDD	0.96
Blasts (%)	%BLASTS	1.8
Polymorphonuclear ratio (%)		29.3
Polymorphonuclear cluster x axis mode	PMNX	64.4
Mononuclear central x channel	MNX	14.7
Mononuclear central y channel	MNY	13.3
Mononuclear polymorphonuclear valley	MNPMN	20

Supplemental Table 2, continued

Hematology Analyzer Data	Abbr.	Value
Red Blood Cell Related		
RBC count (x10^6/μl)	RBC	4.06
Hematocrit (%)	HCT	34.6
Mean corpuscular hemoglobin (MCH; pg)	MCH	30.9
Mean corpuscular hemoglobin conc. (MCHC; g/dl)	MCHC	36.3
RBC hemoglobin concentration mean (CHCM; g/dl)	CHCM	36.7
RBC distribution width (RDW; %)	RDW	14.1
Hemoglobin distribution width (HDW; g/dl)	HDW	2.69
Hemoglobin content distribution width (CHDW; pg)	HCDW	3.50
Normochromic/Normocytic RBC count (x10^6/μl)		340
Macrocytic RBC count (x10^6/μl)	#MACRO	51
Hypochromic RBC count (x10^6/μl)	#HYPO	0.0
Platelet Related		
Plateletcrit (PCT; %)	PCT	0.20
Mean platelet concentration (MPC; g/dl)	MPC	28.9
Platelet conc. distribution width(PCDW; g/dl)	PCDW	5.1
Large platelets (x10^3/μl)	#-L-PLT	4
Platelet clumps (x10^3/μl)	PLT CLU	67

Determining the PEROX Risk Score

With simple modifications to the hematology analyzer (ensuring data export for analysis) and allowing for data entry of clinical and laboratory parameters, calculation of the PEROX risk score can be done in automated fashion. Below is a longhand example.

Step One – Determining whether criteria for each high risk and low risk pattern are met.

Elements used to calculate the PEROX risk score are used by determining in Yes/No fashion whether binary patterns associated with high vs. low risk are satisfied. Elements included in patterns combine a small set of clinical / laboratory data available (age, gender, history of hypertension, current smoking, DM, CVD, SBP, BMI and fasting blood glucose, triglycerides, HDL cholesterol, creatinine, CRP and potassium), combined with data measured during performance of a CBC and differential (not all of these values are reported but they are available within the hematology analyzer).

Supplemental Table 3A below lists the high risk patterns for death and MI. The death high risk pattern #1 consists of a HCDW >3.93 and CHCM <35.07. The example subject has HCDW of 2.69 and CHCM of 36.7. Thus, this subject's data does not satisfy either criterion. Both criteria must be satisfied to have a pattern. This subject therefore does not possess the Death High Risk #1 pattern and is assigned a point value of zero for this pattern. If the subject did fulfill the criterion for the pattern, a point value of one would be assigned.

Death High Risk	Pattern	Subject Values	Pattern	Point Value
1	Hemoglobin content distribution width > 3.93, & RBC hemoglobin concentration mean < 35.07	HCDW=3.50 CHCM=36.7	No	0

The above approach is used to fill in whether each High and Low Risk Patterns are satisfied.

Supplemental Table 3A, indicating whether criteria for each high risk pattern for death and MI are met in this example patient

Death High Risk	Pattern	Subject Values	Pattern Present	Point Value
1	Hemoglobin content distribution width > 3.93, & RBC hemoglobin concentration mean < 35.07	HCDW=3.50 CHCM=36.7	No	0

2	Hypochromic RBC count > 189,
	& Hemoglobin content distribution width > 3.93
		#HYPO=0
		HCDW=3.50
			No	0

3	Mean corpuscular hemoglobin concentration < 34.38, & Perox d/D < 0.89	MCHC=36.3 PXDD=0.96	No	0
4	Hypochromic RBC count > 189, & Macrocytic RBC count > 192	#HYPO=0 #MACRO=51	No	0
5	Mean corpuscular hemoglobin concentration < 33.00, & Mononuclear central x channel < 14.38	MCHC=36.3 MNX=14.7	No	0
6	Age > 67, & Hematocrit < 36.45	AGE=62 HCT=34.6	No	0
7	Mononuclear polymorphonuclear valley < 18.50, Peroxidase y sigma > 9.48	MNPMN=20 PXYSIG=6.55	No	0
8	Mononuclear central x channel < 14.38, & Peroxidase y mean > 19.02	MNX=14.7 PXY=19.06	No	0
9	C-reactive protein > 13.75, & History of hypertension	CRP=1.38 HTN=Yes	No	0
MI High Risk	**Pattern**			
1	Mean platelet concentration > 27.89, & Potassium < 3.85	MPC=28.9 K=4.0	No	0
2	Triglycerides < 130, & Age > 76	TGS=161 AGE=62	No	0
3	RBC distribution width > 13.83, & Lymphocyte count > 1.75	RDW=14.1 #LYMPH=2.10	Yes	1

4	Hypochromic RBC count > 56,
	& Diabetes
		#HYPO=0
		DM=NO
			No	0

5	Body mass index < 24.7, & Neutrophil count < 3.58	BMI=29.0 #NEUT=4.53	No	0
6	Systolic blood pressure > 150, & History of Hypertension	SBP=125 HTN=YES	No	0
7	Polymorphonuclear cluster x axis mode > 29.87, & RBC distribution width > 13.22	PMNX=64.4 RDW=14.1	Yes	1
8	Hemoglobin distribution width > 2.69, & Peroxidase y sigma > 8.59	HDW=2.69 PXYSIG=6.55	No	0
9	Platelet concentration distribution width < 5.39, & RBC hemoglobin concentration mean < 34.69	PCDW=5.1 CHCM=36.7	No	0
10	Mean corpuscular hemoglobin > 32.60, & Male	MCH=30.9 MALE = No	No	0

11	Lymphocyte count < 0.96,
	& Potassium > 4.4
		#LYMPH=2.10
		K=4.0
			No	0

12	Platelet concentration distribution width > 6.04, & Monocyte count > 0.46	PCDW=5.1 #MONO=0.37	No	0
13	Neutrophil cluster mean y< 71.19, & Current smoker	NEUT Y=74.8 SMOKE=No	No	0
14	Mean platelet concentration > 23.19, & Basophil count > 0.12	MPC=28.9 #BASO=0.02	No	0

Supplemental Table 3B, indicating whether criteria for each low risk pattern for death and MI are met in this example patient

Death Low Risk	Pattern	Subject Values	Pattern Present	Point Value
1	RBC hemoglobin concentration mean > 35.07, & Hematocrit > 42.25	CHCM=36.7 HCT = 34.6	No	0

2	Macrocytic RBC count < 192,
	& Age < 67
		#MACRO=51
		AGE=62
			Yes	1

3	RBC hemoglobin concentration mean > 35.07, & RBC count > 4.42	CHCM=36.7 RBC=4.06	No	0
4	Mean platelet concentration > 27.52, & Age < 67	MPC=28.9 AGE=62	Yes	1
5	Peroxidase y sigma < 8.10, & Age < 67	PXYSIG=6.55 AGE=62	Yes	1
6	C-reactive protein < 4.0, & Hematocrit > 42.25	CRP=1.38 HCT=34.6	No	0
7	Hematocrit > 42.25, & Perox d/D > 0.89	HCT=34.6 PXDD=0.96	No	0
8	Mononuclear polymorphonuclear valley > 18.50, & Age < 67	MNPMN=20 AGE=62	Yes	1
9	RBC hemoglobin concentration mean > 35.07, & White blood cell count < 5.86	CHCM=36.7 WBC=7.34	No	0

10	Neutrophil count < 3.96,
	& Age < 67
		#NEUT=4.53
		AGE=62
			No	0

MI Low Risk	**Pattern**			
1	History of cardiovascular disease, & RBC distribution width < 13.22	CAD=NO RDW=14.1	No	0
2	Lymphocyte/Large unstained cell threshold < 44.50, & Blasts (%) < 0.51	LUC=50 %BLASTS=1.8	No	0
3	Systolic blood pressure < 134, & Basophil count < 0.03	SBP=125 #BASO=0.02	Yes	1
4	Platelet clumps > 41, & Fasting blood glucose < 92.5	PLT CLU=67 GLUC=95.2	No	0
5	Hgb distribution width < 2.69, & Hypochromic RBC count < 14	HDW=2.69 #HYPO=0.00	No	0
6	Hypochromic RBC count < 14, & Neutrophil count < 5.83	#HYPO=0.00 #NEUT=4.53	Yes	1
7	Mononuclear central x channel < 12.70, & Neutrophil cluster mean y > 69.30	MNX=14.7 NEUTY=74.8	No	0
8	Mononuclear polymorphonuclear valley > 14.50, & Creatinine < 0.75	MNPMN=20 CREAT=0.83	No	0
9	History of cardiovascular disease, & Systolic blood pressure < 134	CAD=NO SBP=125	No	0
10	Number of peroxidase saturated cells < 0.01, & Neutrophil count < 4.69	#PEROX SAT=0 #NEUT=4.53	Yes	1
11	High density lipoprotein cholesterol> 59, & Mean platelet concentration < 28.56	HDL=44 MPC=28.9	No	0
12	Mononuclear central x channel < 12.70, & C-reactive protein < 5.31	MNX=14.7 CRP=1.38	No	0
13	Mononuclear central x channel < 12.70, & Basophil count < 0.07	MNX=14.7 #BASO=0.02	No	0
14	History of cardiovascular disease, & Neutrophil cluster mean x < 66.07	CAD=0 NEUTX=64.4	No	0

Step Two – Counting the number of high and low risk patterns that are satisfied.

The next step is to count how many positive and negative patterns are fulfilled.
Each high risk pattern has a value of +1 and each low risk pattern has a value of -1.

In our example:

Number of high risk patterns Subject has = 2
Number of low risk patterns Subject has = 7

Step Three – Calculating the weighted Raw Score.

Subjects almost always have combinations of both high and low risk patterns.

Overall risk is calculated by a weighted sum of the number of high risk and low risk patterns.

The weight for each positive pattern is [+1/number of high risk patterns satisfied], while for each negative pattern is [-1/number of low risk patterns satisfied].

Total possible number of high risk patterns is 23.
Total possible number of low risk patterns is 24.

Thus, if a subject had all 23 positive risk patterns and no low risk patterns they would have a maximal Raw Score of +1. If a subject had no high risk patterns and all low risk patterns, they would have a minimum Raw Score of -1.

The Raw Score of a subject is calculated by the weighted sum of high risk and low risk patterns. In this example, we know:

Raw Score = (1/23 possible high-risk patterns) x (number of high-risk patterns satisfied) +
(-1/24 possible low-risk patterns) x (number of low-risk patterns satisfied)

= 1/23 x 2 + -1/24 x 7 = - 0.2047

Note – the Raw Score can have a positive or negative value.

Step Four – Calculating the final PEROX Risk Score

The calculated Raw Score ranges from -1 to +1 with 0 as the midpoint.

The PEROX Risk Score adjusts the range from ±1 to a range of 0 to 100 by assuming 50 (rather than 0) as the midpoint of the scale. This is achieved by multiplying the Raw Score by 50, and then adding 50.

PEROX Risk Score = (50 x Raw Score) + 50
= (50 x - 0.2047) + 50
= 39.8

Figure 1F of the manuscript allows one to use the Perox Risk Score to estimate overall incident risk of death or MI over the ensuing one-year period. In this example, the subject's 1 yr event rate is approximately 2%.

VI. PEROX Model Validation

The Somers' D rank correlation, Dxy, provides an estimate of the rank correlation of the observed binary response and a continuous variable. Thus, it can be used as an indicator of model fit for the PEROX model. Dxy in the PEROX model measures a correlation between the predicted PEROX score and observed binary response (event vs. non-event). We calculated Dxy for both Derivation and Validation cohorts. A large difference in Dxy values between these two cohorts indicate a large prediction error. As can be seen from the table below, there is no evidence of lack of fit since the differences are small for all three cases. Based upon these analyses, the PEROX risk score showed small overall prediction errors (e.g. 3.8% difference between Derivation and Validation Cohorts for one year Death or MI outcome).

Supplemental Table 4:
Model validation of the PEROX model using Dxy

Dxy	Derivation	Validation	Difference (%)
Death	0.607	0.676	11.4
MI	0.319	0.306	4.1
Death/MI	0.501	0.520	3.8

Hosmer-Lemeshow statistic is a goodness of fit measure for binary outcome models when the prediction is a probability. However the PEROX risk score is not a probability, hence the Hosmer-Lemeshow statistic can not be directly applied to PEROX score. We therefore converted PEROX risk scores on a probability scale through a logistic regression model. Then Hosmer-Lemeshow test was applied to examine the goodness of fit using PEROX score as a risk factor for event prediction. As can be seen from the results below, no evidence of lack of fit was observed since all p-values are significantly larger than 0.05.

Supplemental Table 5:
Model validation of the PEROX model using Hosmer Lemeshow test

	χ^2	p-value
Death	8.08	0.426
MI	2.73	0.950
Death/MI	11.68	0.166

To provide further realistic simulation, the method used for generating the PEROX risk score was cross-validated by using ten random 10-folding experiments within the learning dataset (Derivation Cohort). *k*-folding is a cross-validation technique in which the samples are randomly divided into *k* parts, 1 part is used as the test set and the remaining *k*-1 parts are used for training. The test set is permuted by leaving out a different test set each time. In this case, *k*=10 was used and the entire procedure was repeated 10 times, resulting in 100 experiments within the Derivation cohort. The data contains a relatively small proportion of deaths and MIs in 1 year. To ensure that there was a fair sampling of the Death and MI events in all the k-folds, random stratified sampling was performed (meaning that Death, MI, and controls were randomly divided into k parts separately within the Derivation cohort). Within each fold, separate LAD models were built for Death vs. controls and MI vs. controls. Cut-points were selected on the

training data using 3 equal frequency cuts. The Death and MI models were combined and used to compute the PEROX score on the test set. Area under the ROC curve was computed on the test set. The summary results for the 100 experiments are presented in the Table below.

Supplemental Table 6:
Model validation of the PEROX model *k*-folding technique

	25%	50%	75%
AUC	0.68	0.72	0.75

VII. Univariate Cox Proportional Hazard Analysis for Prediction of One-Year Outcomes Using Peroxidase-based Hematology Parameters Included in PEROX Model

	Derivation Cohort	Validation Cohort	Death 1 Year HR (95% CI)	MI I Year HR (95% CI)
White Blood Cell Related				
White blood cell count ($x10^3/\mu l$)	6.50 ± 2.19	6.51 ± 2.22	1.31 (1.21-1.42) *	1.04 (0.91-1.20)
Neutrophil count ($x10^3/\mu l$)	4.39 ± 1.97	4.42 ± 1.94	1.37 (1.26-1.48) *	1.01 (0.88-1.16)
Lymphocyte count ($x10^3/\mu l$)	1.54 ± 0.76	1.52 ± 0.86	0.73 (0.62-0.86) *	1.02 (0.89-1.16)
Monocyte count ($x10^3/\mu l$)	0.35 ± 0.18	0.35 ± 0.17	1.13 (1.09-1.16) *	1.06 (0.96-1.16)
Eosinophil count ($x10^3/\mu l$)	0.21 ± 0.15	0.21± 0.18	1.11 (1.03-1.19) *	1.05 (0.93-1.18)
Basophil count ($x10^3/\mu l$)	0.05 ± 0.03	0.05 ± 0.03	1.09 (0.98-1.21)	1.07 (0.94-1.22)
Number of peroxidase saturated cells ($x10^3/\mu l$)	0.82 (0.30-1.53)	0.80 (0.30-1.50)	1.00 (0.89-1.12)	1.06 (0.91-1.23)
Neutrophil cluster mean x	61.7 ± 6.0	61.7 ± 6.3	0.96 (0.86-1.06)	0.97 (0.85-1.11)
Neutrophil cluster mean y	70.0 ± 6.0	70.0 ± 6.4	1.01 (0.90-1.14)	0.95 (0.84-1.07)
Ky	97.36 ± 2.38	97.25 ± 2.41	0.97 (0.86-1.09) *	0.90 (0.78-1.04)
Peroxidase x sigma	0.01 ± 0.12	0.01 ± 0.12	1.10 (1.03-1.18) *	1.06 (0.96-1.18)
Peroxidase y mean	18.1 ± 0.7	18.1 ± 0.7	1.61 (1.46-1.77) *	1.10 (0.96-1.27)
Peroxidase y sigma	8.11 ± 1.07	8.12 ± 1.05	1.79 (1.61-1.99) *	1.16 (1.01-1.33) *
Lobularity index	1.9 (1.0-2.1)	1.9 (1.0-2.1)	0.92 (0.83-1.01)	1.03 (0.89-1.20)
Lymphocyte/large unstained cell threshold	45.0 ± 1.6	45.1 ± 1.6	1.16 (1.08-1.24) *	1.07 (1.00-1.17)
Perox d/D	0.9 (0.9-1.0)	0.9 (0.9-1.0)	0.91 (0.85-0.97) *	1.16 (0.85-1.56)
Blasts (%)	0.77 ± 0.49	0.77 ± 0.49	1.34 (1.22-1.47) *	1.07 (0.93-1.23)
Polymorphonuclear ratio (%)	1.0 (0.99-1.0)	1.0 (0.99-1.0)	0.77 (0.65-0.90) *	0.99 (0.84-1.15)
Polymorphonuclear cluster x axis mode	27.5 ± 3.6	27.4 ± 3.7	0.91 (0.82-1.02)	1.08 (0.93-1.25)
Mononuclear central x channel	14.1 (13.0-15.0)	14.1 (13.0-15.0)	0.80 (0.74-0.88) *	1.12 (0.95-1.32)
Mononuclear central y channel	14.5 ± 1.1	14.5 ± 1.1	0.79 (0.73-0.87) *	1.04 (0.89-1.20)
Mononuclear polymorphonuclear valley	18.0 (18.0-20.0)	18.0 (18.0-20.0)	0.69 (0.61-0.77) *	1.06 (0.94-1.21)
Red Blood Cell Related				
RBC count ($x10^6/\mu l$)	4.30 ± 0.52	4.33 ± 0.52	0.59 (0.53-0.66) *	0.93 (0.81-1.08)
Hematocrit (%)	40.9 ± 6.2	41.0 ± 4.2	0.51 (0.45-0.59) *	0.78 (0.65-0.93) *
Mean corpuscular hgb (MCH; pg)	30.4 ± 2.1	30.3 ± 2.0	0.83 (0.75-0.92) *	1.03 (0.89-1.19)
Mean corpuscular hgb conc. (MCHC; g/dl)	33.4 ± 5.7	33.4 ± 5.7	0.86 (0.80-0.92) *	0.91 (0.82-1.01)
RBC hgb concentration mean (CHCM; g/dl)	35.1 ± 1.3	35.2 ± 1.3	0.53 (0.49-0.59) *	0.90 (0.78-1.04)
RBC distribution width (RDW; %)	13.4 ± 1.2	13.4 ± 1.2	1.48 (1.42-1.55) *	1.26 (1.14-1.40) *
Hgb distribution width (HDW; g/dl)	2.7 ± 0.3	2.7 ± 0.3	1.52 (1.39-1.66) *	1.26 (1.12-1.43) *
Hgb content distribution width (CHDW; pg)	3.8 ± 0.4	3.8 ± 0.4	1.44 (1.37-1.51) *	1.19 (1.07-1.33) *
Normochromic/Normocytic RBC count ($x10^6/\mu l$)	3.65 ± 0.39	3.66 ± 0.39	0.64 (0.60-0.68) *	0.89 (0.78-1.01)
Macrocytic RBC count ($x10^6/\mu l$)	0.01 (.01-.03)	0.01 (.01-.03)	1.76 (1.55-2.00) *	1.03 (0.89-1.20)
Hypochromic RBC count ($x10^6/\mu l$)	0.006 (0.001-0.002)	0.005 (0.001-0.002)	1.12 (0.99-1.27)	1.18 (1.00-1.38)
Platelet Related				
Plateletcrit (PCT; %)	0.18 ± 0.05	0.18 ± 0.06	1.15 (1.04-1.27) *	0.99 (0.85-1.14)
Mean platelet concentration (MPC; g/dl)	27.1 ± 1.7	27.0 ± 1.7	0.75 (0.68-0.83) *	0.97 (0.84-1.12)
Platelet conc. distribution width (PCDW; g/dl)	5.6 ± 0.4	5.7 ± 0.4	0.95 (0.84-1.06)	0.95 (0.83-1.01)
Large platelets ($x10^3/\mu l$)	4 (3-6)	4 (3-6)	1.10 (0.94-1.28)	1.10 (0.91-1.34)
Platelet clumps ($x10^3/\mu l$)	41.5 ± 37.1	42.4 ± 36.1	1.00 (1.00-1.00)	1.00 (1.00-1.00)

All variables listed were present in the PEROX risk score model. Data are shown as mean ± standard deviation for normally distributed continuous variables, or median (interquartile range) for non-normally distributed continuous variables. Some variables have no unit of measure associated with them. Median for peroxidase X sigma was zero, therefore, mean is shown. Hazard ratios were calculated per standard deviation (for normally distributed variables). For variables with non-normal distribution, values were log transformed and hazard ratios calculated per log of standard deviation. Variable definitions are available in Supplemental Material. Abbreviations: MI, myocardial infarction; HR, hazard ratio; CI, confidence interval; RBC, red blood cell; Hgb, hemoglobin.

VIII. References

1. Lauer MS, Alexe S, Pothier Snader CE, Blackstone, EH, Ishwaran, H and Hammer PL. Use of the logical analysis of data method for assessing long-term mortality risk after exercise electrocardiography. *Circulation.* Aug 6 2002;106(6):685-690.

2. Crama Y HP, Hammer PL, and Ibaraki T. Cause-effect relationships and partially defined Boolean functions. *Annals of Operations Research.* 1988 1988;16(1):299-326.

3. Boros E HP, Hammer PL and Ibaraki T, et al. Logical analysis of numerical data. *Math Programming.* 1997 1997;79:163-19