Supplemental Material

Data S1.

International Classification of Diseases (ICD) codes and Current Procedural Terminology (CPT) codes used in creating the study population

Definition of noninvasive cardiac testing: 1) A completed ED referral to a KPSC cardiology department with CPT code 200267; or 2) an outpatient visit with CPT code for a stress ECG test (CPT: 93015, 93018).

Due to KPSC adoption of ICD-10 starting October 2015, the use of ICD-9 codes covers the period from June through September 2015, and ICD-10 codes cover the remaining period.

Definition of AMI: ICD-9 4100, 4101, 4102, 410.10, 410.11, 410.12, 410.20, 410.21, 410.22, 410.30, 410.31, 410.32, 410.40, 410.41, 410.42, 410.50, 410.51, 410.52, 410.60, 410.61, 410.62, 410.70, 410.71, 410.72, 410.80, 410.81, 410.82, 410.90, 410.91, 410.92 and ICD-10 I21, I22 codes.

Revascularization and CABG were defined using extensive lists of ICD-9, ICD-10, and CPT codes, which we do not present here for brevity but can provide upon request.

Data S2.

NLP extracted and derived variables from ETT reports

ETT reports include a rich set of diagnostic and prognosis information.[1] We extracted the following variables from ETT reports:

Study protocol defines the speed and inclination at specified time intervals. Different protocols were used based on patients' physical conditions.[1, 2] Study protocols we extracted and grouped into two categories:

- **Standard Bruce:** Bruce

- **Other protocols:** modified Bruce, Astrand, Balke, Cornell, Ellestad, Naughton, manual, etc.

Exercise time is denoted in seconds which measures functional capacity and is one of the most influential prognostic factors.[3, 4]

Blood pressure (BP) is measured before and during the ETT. The blood pressure response to exercise can be used to prognosticate the risk of cardiovascular disease.[2, 3]

Heart rate (HR) variables include resting HR, maximum HR, achieved MPHR (maximal predicted heart rate).[3, 5]

Metabolic equivalents (METS) is a functional capacity measurement calculated based on the speed and grade of the ETT.[6] METS was shown to be a powerful predictor of mortality.[7] The METS extracted is a numerical variable.

ECG ST change is the ST-segment depression information extracted from ETT reports.

Symptoms during exercise and *reasons for stopping exercise* refer to patient-reported symptoms and may have prognostic value for future cardiac events or death.[2, 8, 9]

Symptom was grouped into three categories:

- Abnormal: such as angina, typical angina, chest pressure

- Atypical angina: such as atypical angina, burning chest pain, arrhythmia, atrioventricular block, bradycardia, left /right bundle branch block, tachycardia, ST, and blood pressure abnormality

- Atypical symptoms: such as fatigue, shortness of breath, dizziness, hypertensive response, other atypical symptoms

Reason for stopping was grouped into five categories:

- Endpoint: such as fatigue, reach study endpoint

- Noncardiac: such as exercise intolerance, safety, headache, dizziness, nausea

- BP response: such as abnormal BP responses

- Dyspnea: such as shortness of breath, cough

- Cardiac: such as chest pain, arrhythmia, atrioventricular block, cardiac arrhythmia, left bundle branch block, paroxysmal atrial tachycardia, right bundle branch block, sinus bradycardia, sinus tachycardia, ventricular bigeminy, and other cardiac reasons

Clinician assessment is the overall impression of the ETT stated by the clinician, which was classified into four categories: normal, non-diagnostic, equivocal, and abnormal.

From the extracted variables, we further derived the following variables from ETT reports:

   ETT result is the final result of the ETT synthesized by the NLP algorithm based on clinical assessment and other extracted ETT variables. ETT result was classified into four categories: normal, non-diagnostic, equivocal, and abnormal. The last three categories were also referred to as non-normal in this study.

   ECG result refers to ECG changes suggestive of ischemia[2]. Previous studies showed ischemic ECG to be a strong predictor of cardiac events.[10] The ST-segment depression

information extracted from ETT reports was used to derive the final ECG results by combining with the extracted ECG assessment (Supplemental Method 2). Results were defined as normal, non-diagnostic, and abnormal.

Chronotropic index is a heart rate related variable defined as (maximum HR - resting HR) / ((220 - age) - resting HR))

Chronotropic incompetence is defined as Chronotropic index < 0.8

Blood pressure related variables:

Additional variables such as hypertensive and hypotensive response, low SBP (systolic blood pressure) peak were derived based on the definition described below:

- SBP: systolic blood pressure

- DBP: diastolic blood pressure

- Resting pulse pressure: resting SBP – resting DBP

- Maximum pulse pressure: maximum SBP – maximum DBP

- SBP change: maximum SBP - resting SBP

- Hypertensive (diastolic) response: (maximum DBP - resting DBP) > 10 mm Hg or maximum DBP ≥ 100 mm Hg

- Hypertensive (systolic) response: maximum SBP ≥ 210 mm Hg for men and ≥ 190 mm Hg for women

- Hypertensive response: hypertensive (diastolic) response or hypertensive (diastolic) response

- Hypotensive response: maximum SBP < resting SBP

- Low systolic peak: maximum SBP < 140 mm Hg or (maximum SBP - resting SBP) < 10

Data S3.

Algorithm to derive the result of the treadmill test report

Inputs from NLP extracted variables:

- **Assessment**: Normalized ETT result based on clinician's assessment (normal, abnormal, equivocal, non-diagnostic)

- **MHR**: Maximum heart rate achieved in the test

- **MPHR**: Maximum predicted heart rate achieved in the test (%)

- **ECG**: Final ECG component derived by NLP (normal, abnormal, equivocal)

Output:

- **ETT_final**: Final ETT test result derived by NLP (normal, abnormal, equivocal, non-diagnostic, etc.)

    if MPHR is not found:

        if age and MHR are available:

            $MPHR = MHR * 100 / (220\text{-age})$

    if Assessment is found:

        ETT_final = Assessment

    else if MPHR is found:

        if ($MPHR \geq 85$): **ETT_final** = ECG

        else if ($MPHR < 85$) and (ECG is not normal): **ETT_final** = ECG

        else if ($MPHR < 85$): **ETT_final** = non-diagnostic

Data S4.

Sample list of descriptors used to identify subjective assessment in the ETT reports

Abnormal

abnormal, abn, high risk, positive for, complained, c/o

Equivocal

equivocal, borderline, cannot be ruled out, concerning for, could be considered, intermediate

risk, non-specific, possible, seems to be, suggestive, remain a consideration

Non-diagnostic

non-diagnostic, did not achieve, failure to achieve, inconclusive, not decisive, non dx, not

diagnostic for, submaximal, not performed, aborted, cannot perform, not done, unable to walk,

cancelled, deferred, postponed

Normal

normal, does not meet ischemia criteria, lack of, least likely, low risk, low suspicion, negative,

no evidence of, no stress-induced, non ischemic, unlikely, unremarkable

Data S5.

Sample treadmill test report

TREADMILL EXERCISE STRESS TEST (BRUCE PROTOCOL)

Reason for Test: Chest tightness and felt SOB with numbness both arms and legs.

Resting EKG:SR at 60, 1st degree AVHB.

Target HR:85 %

Max Predicted Heart Rate:123 bpm

Pre-test symptoms

MIN MPH % GRADE HR BP COMMENTS

0 Resting 60 80/52 supine Left arm denied chest sx/denied dizziness; baseline mild head discomfort.

Baseline symptoms

1 1.7 10 81 Denied chest sx/denied dizziness; baseline mild head discomfort.

2 1.7 10 88 " "

3 1.7 10 90 87/53 " " PACs

4 2.5 12 96 " "

During exercise symptoms

5 2.5 12 102 tired/pt requested to slow down; chest pressure

7-8/10; denied dizziness; increased baseline head discomfort.

Recovery symptoms

MIN POST EXERCISE HR BP COMMENTS

1 91 84/41 chest pressure 5-6/10; denied dizziness; baseline mild head discomfort.

3 77 110/55 chest pressure resolving; denied dizziness; baseline mild head discomfort.

6 74 100/54 chest pressure resolving; denied dizziness; baseline mild head discomfort.

9 71 87/52 ==chest pressure resolved==; denied dizziness; baseline mild head discomfort.

INTERPRETATION:

Test Stopped after 6 min. 48 sec. of exercise.

pt requested to slow down, feel tired.

Neighboring sentences

Achieved: 78 % PMHR (9.4 METS)

Peak HR achieved:114 bpm

Peak BP achieved:110 / 55 mmHg

EX capacity: below average

B/P response: baseline low blood pressure w/o SX noted.

CP:chest pressure; denied dizziness; increased baseline head discomfort at peak exercise.

ST changes:

Pre ex - none

During - 0.5 mm horizontal-upslope ST depression at peak exercise

Post ex - return to baseline

Section

Ectopy / Arrhythmia(s):

PAC's

IMPRESSION:

Equivocal TMST study with complaint of chest discomfort at peak exercise but no ischemic ECG change at less than 85% MPHR; better than average cardiovascular performance for age and gender by achieving more than 9 METs denied dizziness; increased baseline head discomfort at peak exercise; baseline low blood pressure w/o SX noted.

NLP-extracted information were highlighted in colors.

Data S6.

Algorithm to impute study protocol variable of the treadmill test report

if (3 >= extime > 0) and (4 >= mets > 0):

protocol = 'standard Bruce'

else if (6 >= extime > 3) and (7 >= mets > 4):

protocol = 'standard Bruce'

else if (9 >= extime > 6) and (10 >= mets > 7):

protocol = 'standard Bruce'

else if (12 >= extime > 9) and (13 >= mets > 10):

protocol = 'standard Bruce'

else if (15 >= extime > 12) and (15 >= mets > 13):

protocol = 'standard Bruce'

else if (18 >= extime > 15) and (18 >= mets > 15):

protocol = 'standard Bruce'

else if (21 >= extime > 18) and (21 >= mets > 18):

protocol = 'standard Bruce'

We only focused on identifying whether the missing protocol is a standard Bruce protocol.

extime: exercise time in minutes

mets: Metabolic equivalents

Data S7.

Algorithm to derive the result of the ECG component of the treadmill test report

Inputs from NLP extracted variables:

- **st_mm**: ST change magnitude in mm

- **st_direction**: ST change direction

- **st_text**: ST change text description

- **ECG_text**: ECG assessment text description

Output:

- **ST**: ST change categorical value (normal, abnormal, equivocal)

- **ECG**: Final ECG component derived by NLP (normal, abnormal, equivocal)

Step 1: Convert st_change to st_cat:

   if st_mm is found:

   if (st_mm ≥ 2 mm): **ST** = abnormal

   else if (st_mm ≥ 1 mm) and (st_direction equal 'downsloping'): **ST** = abnormal

   else if (st_mm ≥ 1 mm) and (st_direction equal 'horizontal'): **ST** = abnormal

   else if (st_mm ≥ 1 mm) and (st_direction equal 'upsloping'): **ST** = equivocal

   else: **ST** = normal

   else if st_text is found: **ST** = st_text

Step 2: Combine ST with ecg_text to derive the final ECG:

   if both **ECG_text** and **ST** were found:

   set **ECG** as one of the more severe results of **ECG_text** and **ST**;

else: set **ECG** to the one which was found

Table S1. ETT variables extracted by NLP.

NLP extracted variables	Value	In the reference standard
Clinician assessment	Equivocal	Yes
Study protocol	Bruce	Yes
Exercise time (sec)	408	Yes
Reason for stopping	fatigue	Yes
Symptom	abnormal	Yes
METS	9.4	Yes
Maximum BP	110/55	Yes
MPHR	78	Yes
Maximum HR	114	Yes
Resting BP	80/52	No*
Resting HR	60	No*
ECG ST change	0.5 mm horizontal	No*

*NLP's accuracy was not formally evaluated on these variables because these variables were not manually extracted in the reference standard. However, we manually verified the NLP results on these variables and confirmed that their accuracies are similar to other variables.

Table S2. ETT variables derived based on NLP extracted information.

NLP derived variables	Value	In the reference standard
ETT result	Non-diagnostic, Equivocal	Yes
ECG result	Non-ischemic	Yes
Chronotropic index	0.64	No*
Resting SBP	80	No*
Resting DBP	52	No*
Resting pulse pressure	28	No*
Maximum SBP	110	No*
Maximum DBP	55	No*
Maximum pulse pressure	55	No*
Hypertensive response	No	No*
SBP change	30	No*
Hypertensive response	No	No*
Hypertensive (diastolic) response	No	No*
Hypertensive (systolic) response	No	No*
Hypotensive response	No	No*
Low systolic pressure peak	Yes	No*

*Some variables were not manually extracted in the reference standard and not formally validated. The majority of them were derived based on the variables that were formally validated (Table 2).

Table S3. Kappa scores between the two physicians on the validation dataset measured by the treadmill test variables.

ETT variables	Kappa (95% CI) Reviewer 1 vs. Reviewer 2	Kappa (95% CI) Reviewer 1 vs. Reference Standard	Kappa (95% CI) Reviewer 2 vs. Reference Standard
Study protocol*	0.66 (0.50-0.82)	0.66 (0.50-0.82)	1.0 (1.0-1.0)
Exercise time	0.79 (0.62-0.97)	0.79 (0.62-0.97)	1.0 (1.0-1.0)
Reasons for stopping*	0.73 (0.58-0.87)	0.77 (0.63- 0.90)	0.95 (0.88-1.0)
Symptom*	0.74 (0.58-0.91)	0.89 (0.78-1.0)	0.83 (0.69-0.97)
Symptom2*	1.0 (1.0-1.0)	1.0 (1.0-1.0)	1.0 (1.0-1.0)
ECG*	0.78 (0.59-0.97)	0.86 (0.70-1.0)	0.90 (0.76-1.0)
METS	0.95 (0.86-1.0)	0.95 (0.86-1.0)	1.0 (1.0-1.0)
Maximum BP	0.88 (0.77-1.0)	0.97 (0.91-1.0)	0.91 (0.81-1.0)
MPHR	0.95 (0.86-1.0)	0.95 (0.86-1.0)	1.0 (1.0-1.0)
Maximum HR	0.92 (0.83-1.0)	0.97 (0.92-1.0)	0.94 (0.87-1.0)
ETT results	0.86 (0.78-0.95)	0.90 (0.82-0.98)	0.95 (0.89-1.0)

*For evaluation purposes, the results of these multicategory variables were dichotomized into two categories:

- Study protocol: standard Bruce protocol vs. other types of study protocols
- Reasons for stopping: target heart rate achieved vs. other reasons
- Symptom: no symptoms vs. (abnormal, atypical angina, atypical symptoms)
- Symptom2: no symptoms vs. abnormal
- ECG: (normal, non-diagnostic) vs. abnormal

ETT = exercise treadmill test; BP = blood pressure; ECG = electrocardiogram; HR = heart rate;

METS = metabolic equivalents; MPHR = maximum predicted heart rate; NLP = natural

language processing; NPV = negative predictive value; PPV = positive predictive value.

Table S4. Troponin values by ETT results.

Troponin values (ng/ml)	Normal	Abnormal	Equivocal	Non-diagnostic	p Value	Total
N (%)	3908 (75)	310 (5.9)	344 (6.6)	652 (12.5)	< 0.0001	5214 (100)
< 0.02	3635 (93)	268 (86.5)	317 (92.2)	568 (87.1)		4788 (91.8)
0.02 – 0.5	271 (6.9)	42 (13.5)	27 (7.8)	83 (12.7)		423 (8.1)
> 0.5	2 (0.1)	0 (0)	0 (0)	1 (0.2)		3 (0.1)

Table S5. Number of conflicted or missing cases for selected variables.

ETT variables	Missed cases N (%)
Study protocol	89 (1.7%)
Exercise time*	135 (2.6%)
Resting BP	434 (8.3%)
Maximum BP	434 (8.3%)
Resting HR†	392 (7.5%)
Maximum HR†	275 (5.3)
MPHR‡	44 (0.8%)
METS§	114 (2.2%)
ECG fining	15 (0.3%)
Reason for stopping	833 (12.4%)
Clinician assessment	126 (2.4%)

* Included cases where the difference between conflicted exercise time is more than 1 minute

† Included cases where the difference between HR is more than 10

‡ Included cases where the difference between MPHR is more than 5

§ Included cases where the difference between METS is more than 1

Maximum HR has more substantial numbers of conflicted cases since HR is often documented multiple times in the ETT reports. It does The NLP algorithm chose the largest value as the final Maximum HR for these conflicted cases.

SUPPLEMENTAL REFERENCES:

1.   Ashley EA, Myers J, Froelicher V. Exercise testing in clinical medicine. *Lancet*.
        2000;356:1592-1597.

2.   Fletcher GF, Ades PA, Kligfield P, Arena R, Balady GJ, Bittner VA, Coke LA, Fleg JL,
        Forman DE, Gerber TC, Gulati M, Madan K, Rhodes J, Thompson PD, Williams MA,
        American Heart Association Exercise CR, Prevention Committee of the Council on
        Clinical Cardiology CoNPA, Metabolism CoC, Stroke N, Council on E, Prevention.
        Exercise standards for testing and training: A scientific statement from the american heart
        association. *Circulation*. 2013;128:873-934.

3.   Miller TD. Exercise treadmill test: Estimating cardiovascular prognosis. *Cleve. Clin. J.
        Med.* 2008;75:424-430.

4.   Blair SN, Kohl HW, 3rd, Barlow CE, Paffenbarger RS, Jr., Gibbons LW, Macera CA.
        Changes in physical fitness and all-cause mortality. A prospective study of healthy and
        unhealthy men. *JAMA*. 1995;273:1093-1098.

5.   Astrand I. Aerobic work capacity in men and women with special reference to age. *Acta
        Physiol. Scand. Suppl.* 1960;49:1-92.

6.   Jette M, Sidney K, Blumchen G. Metabolic equivalents (mets) in exercise testing,
        exercise prescription, and evaluation of functional capacity. *Clin. Cardiol.* 1990;13:555-
        565.

7.   Myers J, Prakash M, Froelicher V, Do D, Partington S, Atwood JE. Exercise capacity and
        mortality among men referred for exercise testing. *N. Engl. J. Med.* 2002;346:793-801.

8.   Abidov A, Rozanski A, Hachamovitch R, Hayes SW, Aboul-Enein F, Cohen I, Friedman JD, Germano G, Berman DS. Prognostic significance of dyspnea in patients referred for cardiac stress testing. *N. Engl. J. Med.* 2005;353:1889-1898.

9.   Amsterdam EA, Kirk JD, Bluemke DA, Diercks D, Farkouh ME, Garvey JL, Kontos MC, McCord J, Miller TD, Morise A, Newby LK, Ruberg FL, Scordo KA, Thompson PD. Testing of low-risk patients presenting to the emergency department with chest pain: A scientific statement from the american heart association. *Circulation*. 2010;122:1756-1776.

10.   Bourque JM, Beller GA. Value of exercise ecg for risk stratification in suspected or known cad in the era of advanced imaging technologies. *JACC Cardiovasc. Imaging*. 2015;8:1309-1321.