Accepted: 26 September 2017
DOI: 10.1002/ajim.22786

RESEARCH ARTICLE

WILEY AMERICAN JOURNAL OF INDUSTRIAL MEDICINE

Assessment of cumulative health risk in the World Trade Center general responder cohort

Ghalib A. Bello PhD[1] | Susan L. Teitelbaum PhD[1] | Roberto G. Lucchini MD[1] | Christopher R. Dasaro MA[1] | Moshe Shapiro MS[1] | Julia R. Kaplan MPH[1] | Michael A. Crane MD[1] | Denise J. Harrison MD[2] | Benjamin J. Luft MD[3] | Jacqueline M. Moline MD, MSc[4] | Iris G. Udasin MD[5] | Andrew C. Todd PhD[1]

[1] Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, New York

[2] Department of Environmental Medicine, Bellevue Hospital Center/New York University School of Medicine, New York, New York

[3] Department of Medicine, Stony Brook University Medical Center, Stony Brook, New York

[4] Department of Occupational Medicine, Epidemiology and Prevention, Hofstra Northwell School of Medicine at Hofstra University, Hempstead, New York

[5] Environmental and Occupational Health Sciences Institute, Robert Wood Johnson Medical Center, Piscataway, New Jersey

Correspondence
Ghalib A. Bello, PhD, Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, Box 1057, New York 10029, NY.
Email: ghalib.bello@mssm.edu

Funding information
Centers for Disease Control and Prevention/National Institute for Occupational Safety and Health (CDC/NIOSH), Grant numbers: 200-2002-00384, U10-OH008216/23/25/32/39/75, 200-2011-39356/61/77/84/85/88

Background: Multiple comorbidities have been reported among rescue/recovery workers responding to the 9/11/2001 WTC disaster. In this study, we developed an index that quantifies the cumulative physiological burden of comorbidities and predicts life expectancy in this cohort.

Methods: A machine learning approach (gradient boosting) was used to model the relationship between mortality and several clinical parameters (laboratory test results, blood pressure, pulmonary function measures). This model was used to construct a risk index, which was validated by assessing its association with a number of health outcomes within the WTC general responder cohort.

Results: The risk index showed significant associations with mortality, self-assessed physical health, and onset of multiple chronic conditions, particularly COPD, hypertension, asthma, and sleep apnea.

Conclusion: As an aggregate of several clinical parameters, this index serves as a cumulative measure of physiological dysregulation and could be utilized as a prognostic indicator of life expectancy and morbidity risk.

KEYWORDS
9/11, Health Risk Index, mortality, physiological dysregulation, World Trade Center

1 | INTRODUCTION/BACKGROUND

Responders to the World Trade Center attacks of September 11, 2001 sustained exposures to numerous hazardous elements produced by the collapse of the WTC towers. In the years following, various health conditions have been documented and studied among the cohorts of responders: New York City Fire Department (FDNY),[1,2] non-firefighter "general" responders,[3] and "survivors."[4–6] As the cohorts age, new studies continue to reveal the development and progression of a range of chronic conditions.[1,6–10] In addition,

multimorbidity,[11–13] the coexistence within an individual of more than one chronic condition, has been observed among a portion of general responders.[14] To date, however, there have not been extensive efforts to evaluate the cumulative burden of comorbidities on responders, nor have there been systematic assessments of multimorbidity risk.

There is a growing consensus that non-infectious diseases (particularly cardiovascular, respiratory, and metabolic) may potentially arise from common underlying causes.[11] This idea is supported by evidence showing that chronic conditions often co-occur within the same individual, often at frequencies that exceed that which would be expected by chance alone.[15,16] Efforts to elucidate the patho-biological mechanisms underpinning these associations have led to the development of a paradigm that views chronic diseases as emergent consequences of the complex network of interactions between various biological systems.[17] From this perspective, chronic diseases are seen as the byproducts of dysregulation in these networks that span various physiological systems.[11]

Recent years have seen increasing adoption of this "systems medicine" perspective in the study of diseases.[18] Rather than examining diseases in isolation, this paradigm advocates a more holistic approach that focuses on assessing and tracking the underlying physiological dysregulation that is thought to give rise to multiple, chronic diseases.[19] Evaluating the extent of multisystem dysregulation can provide key insights into cumulative disease burden and multi-morbidity risk. One popular approach to quantifying multisystem physiological dysregulation is the use of clinical biomarkers which measure the function and integrity of various physiological systems in the body. Aggregating multiple such clinical biomarkers into a composite score provides a measure for assessing the overall physiological state of an individual. The progression from normal/healthy to pathological states is an incremental process, and depending on how far along in this process an individual is, biomarker-based measures of physiological integrity could reflect the cumulative burden of current diseases and/or the risk for future diseases. To this end, multiple studies have documented the usefulness of biomarker-based systemic measures of physiological health such as allostatic load,[20–28] biological age,[29–33] and physiological dysregulation metrics.[34–42] These measures are typically developed by combining multiple, clinical biomarkers of physiological function into a single index. For example, allostatic load, a cumulative risk index of multisystem dysregulation, is commonly computed by counting the number of stress, inflammatory, and metabolic biomarkers for which an individual falls outside the clinically "healthy" range.[21,23,43] Biomarkers typically used for allostatic load include stress indicators (eg, cortisol, catecholamines, dopamine), inflammatory markers (eg, C-reactive protein) and metabolic/cardiovascular function indicators such as cholesterol, blood pressure, and waist-to-hip ratio.[25] Another common cumulative risk index is Biological Age, which is derived by combining information from multiple measures of physiological function into a single, latent variable that can be used to quantify an individual's biological state.[44] Common formulations of Biological Age measures are constructed from clinical biomarkers that span multiple organ systems, such as C-reactive protein, albumin, FEV_1 (Forced Expiratory Volume in 1 sec) and urea nitrogen. The biomarkers are condensed into a single measure of cumulative biological risk via a number of approaches such as principal components analysis[29] and regression onto chronological age using parametric (eg, Levine's method[31,44]) or semi-parametric (eg, Klemera-Doubal method[30]) models. Studies have shown that, compared to chronological age, these Biological Age measures are more precise estimates of an individual's physiological health and remaining life expectancy.

One particularly useful class of cumulative risk indices are mortality risk scores.[45,46] They are typically developed by regressing multiple biomarkers onto a mortality outcome such as a binary indicator of 5-year mortality, or continuous censored survival time. Predictive or regression models/algorithms estimate the underlying functional relationship between a panel of biomarkers of physiological function and mortality, and the resulting association model is then used to aggregate these clinical biomarkers into a unidimensional construct representing mortality risk. Perhaps unsurprisingly, the indices constructed via this approach tend to be stronger predictors of mortality than other classes of cumulative risk indices (eg, allostatic load scores), but an additional advantage is that since mortality is such a strong proxy measure of underlying physiological dysregulation (and indeed its ultimate byproduct), mortality risk indices robustly reflect the extent of multisystem physiological "wear and tear" in an individual. Studies have provided empirical support for this idea, showing that mortality-calibrated indices can predict various chronic disease endpoints.[46–51] There is an intuitive, direct relationship between individuals' overall physiological state of health and their life expectancy, making mortality risk scores potentially useful as a heuristic gauge of overall physiological integrity.

With several years of biomarker and health data collected on a substantial fraction of the WTC General Responder Cohort (GRC), we have the opportunity to begin examining and characterizing multi-systemic physiological dysregulation among this cohort. We propose an integrative approach to examining the overall disease burden and risk within the WTC GRC by developing an index of multisystem biological risk. Following the previously described approaches, we construct a multi-biomarker index using markers from various physiological systems: metabolic, cardiovascular, respiratory, kidney, liver, and immune function (see Table 1). We do this by modeling mortality as a function of all the clinical biomarkers. The resulting model is then used as a formula for combining the biomarker measurements into an index associated with life expectancy. The expectation is that this index is a measure of physiological dysregulation, as it represents a model-driven aggregation of the collective health information from all the biomarkers. We confirm this by testing its association with mortality (in an independent cohort) and assessing its correlation with self-assessed health and a number of disease endpoints.

2 | MATERIALS AND METHODS

Following the previously described approaches, we constructed a multi-biomarker index using markers from various physiological systems: metabolic, cardiovascular, respiratory, kidney, liver, and

TABLE 1 List of variables used to construct the risk index

Variable	Physiological system
Systolic blood pressure Diastolic blood pressure	Cardiac
Body mass index Serum glucose Chloride Calcium Potassium Triglycerides Total cholesterol	Metabolic
Creatinine Blood urea nitrogen	Kidney
White blood cell count Albumin Globulin	Immune
Alkaline phosphatase Aspartate Aminotransferase (AST) Alanine Aminotransferase (ALT) Bilirubin	Liver
Forced Expiratory Volume in one second (FEV_1) (% predicted)* Forced Vital Capacity (FVC) (% predicted)* Platelet count	Respiratory
Red cell distribution width Mean platelet volume Red blood cell count Hemoglobin Mean corpuscular volume Mean corpuscular hemoglobin Mean corpuscular hemoglobin conc	Hematologic

immune function (see Table 1). Using mortality as a validation criterion, we derive an index that aggregates the collective health information from these biomarkers. We confirm that this index is associated with mortality, and assess its association with various health and disease endpoints.

2.1 | Index development

2.1.1 | Training data

The index was developed using data from the third NHANES (1994),[52] conducted by the National Center for Health Statistics (NCHS) from 1988 to 1994. NHANES was designed to assess the health and nutritional status of the US population. Multistage sampling was used for random selection from the non-institutionalized US population. Our analysis utilized the data from NHANES III that was also available for the WTC GRC (routine blood chemistry tests, spirometry, and anthropometric measures—see Table 1). This set of clinical parameters was chosen based on relevance to physiological function and integrity of various organ systems, and also because data on them is available in both the NHANES and WTC general responder cohort records. Data on the survival status of NHANES III participants (as of December 2011) was obtained from the NHANES III Linked Mortality files[53] that are the result of efforts by the NCHS to conduct mortality linkage of NHANES participants to death certificate data available in the National Death Index.

2.1.2 | Predictive modeling

The Gradient Boosting Machine algorithm[54] was used to train a risk model by utilizing clinical chemistry lab test, pulmonary function, and physical examination data from NHANES III. Gradient Boosting is a popular machine learning technique used for nonparametric estimation of the association between a set of explanatory variables and a response of interest. It implements functional gradient descent (a numerical optimization technique) to create an aggregate model based on a sequence of simpler models (referred to as "base learners"). Each of these base learners is designed to iteratively reduce the variation in the response that is not yet explained by the model.[55] For our study, we used an adaptation of this algorithm developed for analysis of right-censored survival outcomes. The base learners used in this study were a sequence of regression trees aggregated to produce a nonparametric model of the relationship between the clinical variables and mortality. In total, 40 000 trees were used, each with a depth of 4. Depth, in this context, is a parameter relating to the number of recursive splits in each tree, with higher depth corresponding to greater complexity in tree architecture. A shrinkage parameter of 5×10^{-4} was chosen to control the learning rate, that is, the magnitude of contribution each base learner makes to the final model. The resulting ensemble of trees formed a risk model used to produce the index, a numeric score representing mortality risk or overall physiological health.

2.2 | Index validation/testing

2.2.1 | Validation data: WTC general responder cohort

The WTC GRC is a large cohort of general (principally non-firefighter) responders. Details on this cohort have already been described elsewhere (see Dasaro et al. (2016)[3]). Briefly, enrollment to the WTC Health Program (WTCHP) requires meeting the following eligibility criteria: 1) individual handled and/or processed human remains as a member of the New York City Office of Chief Medical Examiner; 2) worked for the Port Authority Trans Hudson (PATH) Corporation and during the period between February and July 2002, spent more than 24 h cleaning PATH tunnels; 3) participated in the WTC cleanup effort for at least 4 hours anytime during September 11-14, 2001, or for at least 24 h in the month of September 2001, or for a total of at least 80 h in the 10 months following the attacks.[3]

Individuals enrolled in the WTCHP undergo health monitoring through periodic visits to WTCHP clinics. These monitoring visits were originally scheduled at 18-month intervals but subsequently became annual.[3] At each monitoring visit, members receive a battery of clinical tests, questionnaires, and interviews. Full details of these assessments are given in Dasaro et al[3] but we provide a brief overview here. Physical examinations are performed at every monitoring visit by

qualified physicians. These exams are similar to routine clinical physicals and involve measurement of body measures like blood pressure, height, weight, pulse rate, etc. In addition, clinical chemistry laboratory tests are performed on blood samples drawn during the exam.[3] These blood tests include the Comprehensive Metabolic Panel (CMP),[56] Lipid Profile, Complete Blood Count (CBC),[57] and urinalysis. Pulmonary function is also assessed at every visit by trained personnel.

At Visit 1 only, members complete the Exposure Assessment Questionnaire (EAQ) which is designed to evaluate the extent of exposure to pollutants prior to, and while working on the rescue and recovery effort. For our analyses, we used responses from the EAQ to assess the extent of exposure, using a procedure outlined in Webber et al.[14] This yielded an exposure severity variable with four levels: low, intermediate, high, and very high exposure. These categories were defined based on a cohort member's duration of work on the WTC cleanup effort, exposure to the dust cloud on 9/11, and whether or not they worked on the debris pile (complete details are available in Webber et al[14]). In this study, we merged the "high" and "very high" exposure categories into one category.

At Visit 1 and subsequent visits, a structured, medical-history questionnaire is administered to each GRC member by a trained medical professional. This questionnaire (Interviewer-Administered Medical Questionnaire [IAMQ]) gathers information such as demographics, history of alcohol and tobacco use, employment status and medical conditions (self-reported). In this study, information from this questionnaire was used to identify the occurrence of the following 17 conditions/illnesses among the cohort: acute bronchitis, asbestosis, asthma, chronic obstructive pulmonary disease (COPD), Crohn's disease, gastroesophageal reflux disease (GERD), hepatitis, hypertension, pleurisy, pneumonia, renal failure, rhinitis, sinusitis, sleep apnea, major stroke (excluding transient ischemic attacks), tuberculosis and thyroid disease. In our study, we assessed occurrence of these conditions solely by self-report, specifically via the responses of GRC members to IAMQ items inquiring if and when they had been diagnosed with each condition. Beyond self-reports, we did not use any other form of ascertainment, for example, physician certification.

Beginning from Visit 1, members complete the Self-Administered Mental Health Questionnaire (SAMHQ). This questionnaire solicits, among other things, information about general wellbeing, which is assessed using items from the SF-12©, a validated survey designed for self-assessment of physical and mental health. In this study, we used the Physical Composite Score (PCS), aggregated from a set of questionnaire items that capture the respondent's assessment of their physical health and sense of wellbeing. Because access to SF-12 measures from Visit 1 was unavailable, we used Visit 2 PCS measures in this study.

Mortality data was obtained via National Death Index (NDI) linkage. For a portion of the cohort, their identifying information (names, date of birth, social security number, etc.) was provided to the NDI and probabilistic matching[58] was carried out to determine occurrence of death through December 31, 2011. For the remainder (those with WTCHP follow-up visits after 2011), their identifying information was not sent to NDI because their post-2011 follow-up visits confirmed they were alive beyond December 31, 2011.

As of December 31 2015, 52 492 individuals have sought enrollment into the WTCHP. Out of this, 35 534 individuals met the eligibility criteria, completed a first visit and consented to have their data aggregated. Records on these 35 534 individuals are accessible via the WTCHP General Responder Data Center. However, complete monitoring visit data on some individuals was not available in the database. In particular, records on clinical chemistry laboratory test results were unavailable for a subset of these 35 534 GRC members. Availability of laboratory exam data largely depends on which WTCHP clinic conducted a member's monitoring visit: initially, all lab tests were performed by the same clinical chemistry laboratory and were available to the Data Center. Subsequently; however, clinics were free to use other labs and forwarding the lab data to the Data Center was optional. Labs results were, however, available for monitoring visits held at the largest WTCHP clinic (Mount Sinai Selikoff Center for Occupational Health).[3] Because the risk index developed and tested in this study is composed mostly of clinical parameters from laboratory exams, GRC members with no laboratory data were excluded from the study, leaving 22 833 individuals with at least 10 of the 25 lab exam biomarkers used in the index. As depicted in Figure 1 below, all the analyses we performed to validate the index were carried out on subsets of this sample of 22 833 GRC members. The following sections describe the computation of the index among this sample and the subsequent validation analyses.

2.2.2 | Computation of index in validation samples

The gradient-boosted survival model, described above, uses NHANES III data to generate a complex nonparametric model of the functional relationship between clinical biomarkers and mortality. Given a set of clinical biomarkers, the model outputs an estimate of mortality risk. Missing values were handled automatically by the gradient boosting implementation used in our analysis, which uses a tree-based strategy to generate surrogate estimates for cases with an incomplete set of biomarkers.[59] This risk model was applied to the 22 833 WTC GRC members (see Figure 1) available for analysis. For each individual, we input their clinical biomarker measures (laboratory tests, spirometry, and anthropometric measures [BMI and blood pressure]) into the model to generate a mortality risk estimate. Note that the measured values of the clinical biomarkers are directly input into the model without any prior conversion. Unlike allostatic load computation wherein clinical reference intervals are used to convert a biomarker measure to a binary indicator of whether the measured value falls within or outside "normal" ranges, our approach uses the actual measured value of the biomarker. With the resulting mortality risk estimate, an index score was derived using a simple transformation—adding a constant offset to the estimate and then scaling by a small positive constant. The offset is added so that the resulting index scores are non-zero and positive. Higher values of the index indicate greater health risks and there is no upper bound. To aid in interpretation, the raw index score can be discretized using pre-defined cut-points to

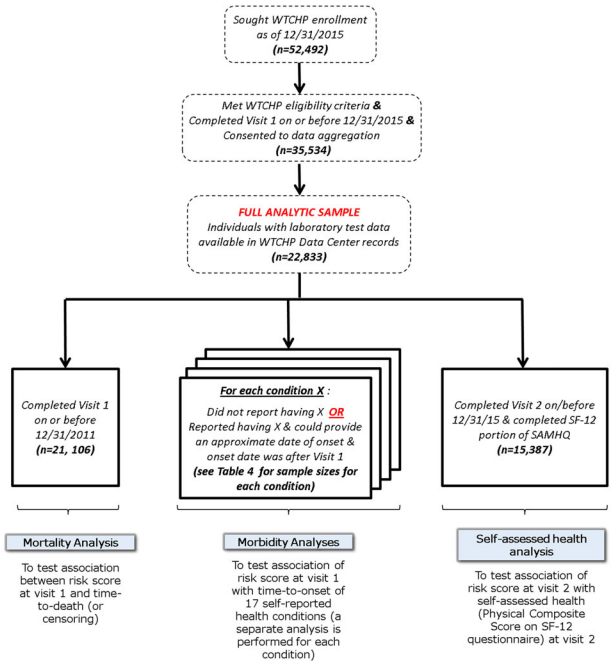

FIGURE 1 Flow diagram showing sequence of selection steps leading to the final sample used in this study (n = 22 833). All validation analyses (summarized at the bottom of the diagram) were based on subsets of the full analytic sample (n = 22 833). For the mortality analysis, those who enrolled (and completed Visit 1) after 12/31/11 were excluded since mortality status information was unavailable for them, leaving n = 21 106. For the morbidity analyses, each condition was analyzed separately. For each, those reporting disease onset prior to Visit 1 were excluded (see Table 4 for final sample sizes for each condition). Because we did not have data on self-assessed health (SF-12 PCS) records at Visit 1, analysis of the relationship between this measure and the risk score was conducted based on measures taken at Visit 2

enable stratification of individuals into risk groups (see section 3). Alternatively, an individual's raw risk score can be compared to age- and sex-specific reference curves, to determine the percentage of individuals of similar age and gender whose raw risk scores fall below that of the individual in question. Using this scheme, an individual with a higher percentile would be interpreted as one whose risk is elevated compared to members of their age/gender cohort.

2.2.3 | Predictive accuracy of risk index

After computation of the risk index score for the entire analytic sample, we performed a number of validation analyses on subsets of this sample. Figure 1 summarizes the series of analyses performed and enumerates the subsets of the full analytical sample used for each analysis.

1) Association of risk scores with mortality:

A Cox proportional hazards model was used to test the association of the Visit 1 risk scores with eventual mortality among the WTC GRC (adjusting for potential confounders). As discussed earlier, mortality status as of 12/31/2011 was ascertained via NDI matching. But for those who enrolled (and completed Visit 1) after this date, information on their mortality status is not yet available. Therefore, the mortality analysis was limited to individuals who had completed Visit 1 by 12/31/2011. Out of the 22 833 individuals selected for this study, around 8% (n = 1727) did not complete Visit 1 until after 2011 and were, therefore, excluded from the mortality analysis (see Figure 1). In the Cox proportional hazards model, survival time was defined as the total time elapsed from the date of Visit 1 to the date of death or, in censored cases (those with no record of death as of 12/31/2011), the last point in time at which they were known to be alive. For censored cases who never returned for a subsequent monitoring visit after 12/31/2011, the last point in time at which they were known to be alive was defined as 12/31/2011. On the other hand, for censored cases who completed at least one subsequent monitoring visit after 12/31/2011, the date of their most recent visit was chosen as the last point in time at which they were known to be alive.

In the Cox proportional hazards model, the risk score was included as a predictor/explanatory variable, along with the following covariates: age at Visit 1, sex, race/ethnicity (White, Black, Hispanic, Other), program entry date (2001-2005, 2006-2008, 2009-2011), 9/11 exposure severity (low, medium, high/very high),[14] smoking status at Visit 1 (never, former, or current) and pre-9/11 occupation (divided into the following occupation groups: construction, protective services, maintenance, other). Occupations falling under the maintenance category are those involving installation, maintenance or repair (electrical, telecommunications, and others). These four occupational categories were chosen because they reflect a responder's role in the rescue and recovery efforts.[60]

2) Association of risk score with self-reported conditions:

Using Cox proportional hazards models, we assessed the prospective association of the index score (computed at Visit 1) with subsequent development of the following 17 health conditions: acute bronchitis, asbestosis, asthma, chronic obstructive pulmonary disease (COPD), Crohn's disease, gastroesophageal reflux disease (GERD), hepatitis, hypertension, pleurisy, pneumonia, renal failure, rhinitis, sinusitis, sleep apnea, major stroke (excluding transient ischemic attacks), tuberculosis, and thyroid disease. For each of these conditions, a separate Cox model was fitted wherein survival time was defined as the amount of time elapsed from Visit 1 date to the date of diagnosis or onset of the condition. In each model, the risk index was included as a predictor, along with the following covariates: age at Visit 1, sex, race/ethnicity (White, Black, Hispanic, Other), program entry date (2001-2005, 2006-2008, 2009, and beyond), 9/11 exposure severity (low, medium, high/very high),[14] smoking status at Visit 1 (never, former, or current) and pre-9/11 occupation (construction, protective services, maintenance, other).

For each condition, occurrence and date of onset/diagnosis were determined solely via self-report, based on responses provided in the IAMQ. The Cox model for each condition was fitted using only individuals who either never reported having the condition, or who did report having the condition and met both of the following criteria: (I) they could provide an approximate date of onset (see below for more details) and (II) this date of onset was subsequent to Visit 1 (this was

done to guarantee a prospective design). Additionally, in the Cox model for hypertension, we included a third criterion: blood pressure measured at the Visit 1 physical exam had to be below hypertensive levels, that is, those with systolic BP ≥ 140 mmHg or diastolic BP ≥ 90 mmHg at baseline (Visit 1) were excluded from the analysis.

For each condition, while some individuals were able to provide exact dates (year, month, and day) of onset, others could not, in which case the following system was used to define a date: if the individual could only recall the diagnosis month and year (but not the exact day), the midpoint of the reported month was used as diagnosis date; if the individual could only recall diagnosis year, the midpoint of the reported year (June 30) was used as the diagnosis date; if the individual could not recall the exact year but could confirm a 2-year range of diagnosis (eg, sometime in 2006-2007), the midpoint of this range was used (eg, for 2006-2007, 12/31/2006 was used as diagnosis date).

Based on these criteria, the final sample sizes varied from condition to condition, that is, different subsets of the full analytic sample of 22 833 GRC members met each condition's selection criteria (see Table 4 for final sample sizes). Since a separate Cox model was fitted for each of the 17 conditions, we adjusted for multiple testing using the Bonferroni procedure. *P* values were deemed significant if they fell below a Bonferroni threshold of 2.94×10^{-3} (=0.05/17).

To assess the association of the risk score with self-reported conditions among various subgroups of the GRC, we performed a series of stratified analyses. The Cox Proportional Hazards models described above were fitted for subjects within each stratum of the following variables: entry year (ie, year of enrollment), age group and time of arrival to the WTC site.

3) Association of risk scores with self-assessed health

Since mortality risk is a strong proxy for overall health, we tested the association of the mortality risk index with self-assessed health status measured using the Physical Composite Score (PCS) of the SF-12, which has been shown to be a valid measure of physical functioning/health.[61,62] The SF-12 is part of the SAMHQ (Self-Administered Mental Health Questionnaire) which was administered starting in Visit 1 for all GRC members. However, we only had data on the SF-12 beginning from Visit 2 onward. Therefore, this cross-sectional analysis tested the association of the risk score (computed from clinical measures taken at Visit 2) with self-assessed health (determined also at Visit 2), adjusting for the following covariates: age at visit 2, sex, race/ethnicity (white, black, Hispanic, other), smoking status at visit 2 (current, former, never), program entry date (2002-2005, 2006-2008, 2009, and beyond), pre-9/11 occupation (construction, protective services, maintenance, other) and exposure severity (low, medium, high/very high).[14] Since the response variable (PCS) in this model did not follow a normal distribution (see section 3), quantile regression[63] was used.

Association of individual clinical parameters with mortality

Using multivariate Cox proportional hazards analysis, we assessed the association between clinical measures (eg, blood count measures, lipid profile, liver function panel) and mortality, adjusting for demographics, program entry date, smoking, pre-9/11 occupation, and 9/11 exposure severity.

Cox proportional hazards models were run using the PHREG procedure in SAS® 9.4 (Cary, NC) and the R package coxph.[64] Quantile regression models were fitted using the SAS® procedure QUANTREG. Gradient boosting was implemented using R package gbm.[64] Chi-squared tests were carried out using the FREQ procedure in SAS.

3 | RESULTS

3.1 | Summary statistics on full analytic sample

Table 2 summarizes demographics of the WTC GRC members included in the full analytic sample (n = 22 833) from which all the validation analyses were derived (see Figure 1). The average age on 9/11/2001 was 39.3 years. The majority of these individuals were male (84.8%), non-Hispanic Caucasian (56.7%), and in protective services (45.7%) or construction-related (24.1%) occupations prior to 9/11. Nearly half of these individuals (49.4%) enrolled in the WTCHP (and completed their first monitoring visit) between 2002 and 2005. At the first visit, a majority of individuals (~59%) reported never having smoked. And based on responses to the Exposure Assessment Questionnaire, 9/11 exposure severity in most of these individuals was ranked either as intermediate (61.7%) or high/very high (20.8%).

Tables S2 and S3 (in the Appendix) present the same demographic summary statistics shown in Table 2, but stratified by age group and entry (enrollment) year. These tables show that the patterns of demographic and exposure characteristics among the cohort vary by age and year of entry.

Table S1 (Appendix) provides summary statistics on measured values (at Visit 1) of all the physiological parameters used in our index. As discussed in the section 2, the measured values of these parameters at Visit 1 were used to compute risk index scores for individuals in the analytic sample. Figure 2 shows the distribution of the Visit 1 risk index scores among members in the full analytic sample. This distribution is right-skewed, reflecting the relatively lower prevalence of extremely high risk scores (Figure 2).

3.2 | Association of risk index with mortality

NDI matching identified 261 deaths among the subset of the full analytic sample with mortality status information (n = 21 106). The median survival time from Visit 1 to date of death was 3.7 years (IQR: 3.98 years). For censored cases, median time from Visit 1 to the last point at which they were known to be alive was 8.9 years (IQR: 5.2 years). Table 3 shows summary statistics for the mortality analysis sample. In this sample, the distributions of age, sex, race, smoking status, pre-9/11 occupation, and exposure severity were similar to those of the full analytic sample of GRC members (see Table 2).

Despite being developed using data from an independent cohort (NHANES III), the value of the risk index computed for GRC members at Visit 1 showed a significant association (P < .0001, HR [95%CI]: 1.347 [1.28, 1.42]) with subsequent mortality after adjustment for

TABLE 2 Summary of demographic and exposure characteristics of the full analytic sample

Characteristics	Summary statistics
Sample size	22 833
Age at 9/11/2001 (mean)	39.3
Age at Visit 1 (mean)	43.6
Risk index at visit 1—Median (quartiles)	14.2 (13.1, 15.8)
Sex	
Female	3473 (15.2%)
Male	19 358 (84.8%)
Missing	2 (0%)
Race/Ethnicity	
White	12 951 (56.7%)
Black	2968 (13%)
Hispanic	5914 (25.9%)
Other	675 (3%)
Missing	325 (1.4%)
Smoking status at Visit 1	
Never	13 346 (58.5%)
Former	5330 (23.3%)
Current	3397 (14.9%)
Missing	760 (3.3%)
Pre-9/11 occupation	
Construction	5493 (24.1%)
Protective	10 446 (45.7%)
CM & IRG*	2037 (8.9%)
Other	4556 (20%)
Missing	301 (1.3%)
Entry year (Visit 1 year)	
2002-2005	11 280 (49.4%)
2006-2008	6712 (29.4%)
>2008	4841 (21.2%)
Exposure severity	
Low	3223 (14.1%)
Intermediate	14 081 (61.7%)
High/Very high	4750 (20.8%)
Missing	779 (3.4%)

*CM, cleaning/maintenance of buildings and grounds; IRG, installation/repair groups (electrical, telecommunications, and others)

demographic covariates, smoking status, entry year, and exposure severity. The results of the Cox proportional hazards analysis are summarized in Table 3. In addition to the risk index, the following variables showed statistically significant associations with mortality: age, smoking status at visit 1, pre-9/11 occupation and WTCHP enrollment year. GRC members who enrolled early in the program had significantly higher mortality risk than those who enrolled later. This effect is likely due to the fact that those with health issues in the first few years after 9/11 were more likely to enroll in the early years of the program than in later years.

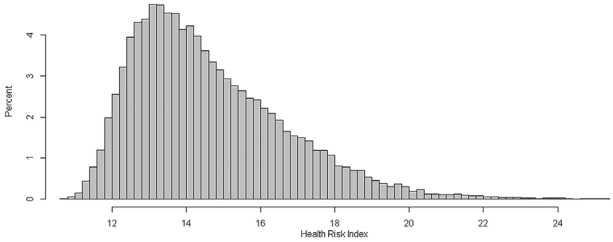

FIGURE 2 Histogram depicting the distribution of risk index scores (computed at Visit 1) across full analytic sample (n = 22 833)

In a secondary analysis, tertiles of the observed index score distribution were used to stratify the sample of GRC members into 3 risk groups: Low (index <13.4), Medium (index: 13.4-15.2), and high risk (index >15.2). To compare survival trends among individuals in these three risk groups, Kaplan-Meier curves plots were generated (shown in Figure 3). These survival curves were adjusted for age and sex using the "direct adjustment" method of Gail and Byar.[65] Comparison of the survival trends revealed a significant overall trend effect ($P < 0.0001$). But while the high risk group showed significantly different survival trends ($P < 0.0001$) from the low and medium risk groups, the latter groups were not significantly different with respect to survival trend ($P = 0.074$).

3.3 | Association of derived index with health conditions

Table 4 summarizes the 17 self-reported conditions examined in this study. For analysis of each condition, the number of individuals meeting the selection criteria (described in the section 2) is given in Table 4. Sample sizes ranged from 12 176 (for sinusitis) to 16 665 (for asbestosis), indicating that all analyses utilized at least ~50% of the full analytic sample (n = 22 833). Table 4 also shows, for each condition, the frequency (number and percentage) of occurrence. Respiratory and cardiovascular/metabolic conditions were the most prevalent. Median time from Visit 1 to disease occurrence ranged from 2.4 years (for asthma) to 5.2 years (for major stroke). The median number of conditions after enrollment was 1, and mean was 1.23.

The risk index was found to have significant associations with multiple conditions. After Bonferroni adjustment to control family wise error rate, the risk index showed statistically significant associations with 11 of the 17 examined conditions. For each condition, the risk index P-value is depicted in Figure 4. The P-values are negative log-transformed so that higher values denote greater significance. The dotted line represents the Bonferroni significance threshold for 17 tests. The strongest associations were observed for hypertension, chronic obstructive pulmonary disease (COPD), asthma, and sleep apnea. The weakest associations were observed for sinusitis, rhinitis and major stroke, which only barely maintained significance after Bonferroni adjustment. For the 11 conditions showing significant associations with the risk index, hazard ratios (and corresponding 95%

TABLE 3 Summary statistics and results of Cox proportional hazards analysis examining association of risk index with mortality (adjusting for covariates)

	Summary statistics	
Variables	***N* = 21106**	***P*-value***
Risk index—Median (quartiles)	14.2 (13.05, 15.77)	<0.0001
Age at visit 1—Mean	43.1	<0.0001
Sex—*N* (%)		0.7827
Female	3276 (15.5%)	
Male	17 828 (84.5%)	
Missing[b]	2 (0%)	
Race/Ethnicity—*N* (%)		0.16
White	11 983 (56.8%)	
Black	2803 (13.3%)	
Hispanic	5649 (26.8%)	
Other	613 (2.9%)	
Missing[b]	58 (0.3%)	
Smoking status at Visit 1—*N* (%)		0.0308
Never	12 346 (58.5%)	
Former	4852 (23%)	
Current	3193 (15.1%)	
Missing[b]	715 (3.4%)	
Pre-9/11 occupation—*N* (%)		0.0054
Construction	5317 (25.2%)	
Protective	9448 (44.8%)	
CM & IRG[a]	1895 (9%)	
Other	4211 (20%)	
Missing[b]	235 (1.1%)	
Entry year—*N* (%)		<0.0001
2002-2005	11 280 (53.4%)	
2006-2008	6712 (31.8%)	
2009-2011	3114 (14.8%)	
Exposure severity—*N* (%)		0.5861
Low	2922 (13.8%)	
Intermediate	13 210 (62.6%)	
High/Very high	4293 (20.3%)	
Missing[b]	681 (3.2%)	

[a]CM, cleaning/maintenance of buildings and grounds; IRG, installation/repair groups (electrical, telecommunications, and others).
[b]Individuals missing values on one or more covariates were not included in the model.

confidence intervals) are summarized in a forest plot (second panel) in Figure 4.

Stratified analyses were performed to determine the association of the risk index with various health conditions within subgroups of the general responder cohort. These subgroups were defined by strata of entry year (2002-2005, 2006-2008, 2009, and beyond) and age group

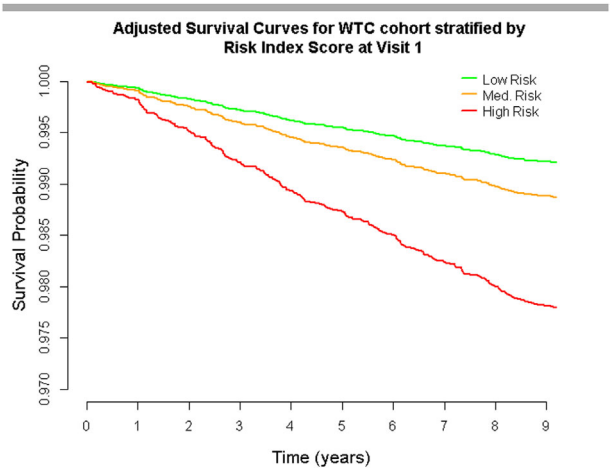

FIGURE 3 Age- and sex-adjusted Kaplan-Meier Curves for risk groups defined by tertiles of the Risk Index (computed from clinical measures taken at Visit 1): Low (<13.4), Medium (13.4-15.2), and high risk (>15.2)

(18-35, 36-50, 50, and older). We also stratified according to time of arrival at the WTC site (9/11/2001-9/13/2001 vs 9/14/2001 or after). We considered responders who arrived on September 11-13 as "first responders," distinct from responders who arrived later (on September 14 or anytime in the subsequent weeks and months).

TABLE 4 Summary of self-reported conditions examined in this study

Self-reported condition	Analytic sample size	# of self-reported diagnoses	% Occurrence after visit 1	Median time from Visit 1 to disease onset* (years)
GERD	13 086	4089	31.20%	2.7
Hypertension	10 297	2328	22.60%	3.5
Rhinitis	14 227	4172	29.30%	2.7
Sinusitis	12 176	3377	27.70%	2.5
Sleep apnea	15 602	2959	19.00%	4.1
Asthma	13 702	2131	15.60%	2.4
Acute bronchitis	12 181	1817	14.90%	3.2
Renal failure	14 697	1132	7.70%	4
COPD	16 221	1171	7.22%	3.6
Pneumonia	14 258	1011	7.10%	4
Thyroid	14 951	976	6.50%	4.7
Major stroke	15 400	320	2.10%	5.2
Hepatitis	14 905	228	1.50%	4.6
Crohn's disease	15 338	216	1.40%	4.5
Asbestosis	16 665	213	1.30%	4.6
Pleurisy	16 347	215	1.30%	3.9
Tuberculosis	16 517	194	1.20%	3.6

*Type III *P*-values from Cox proportional hazards regression.

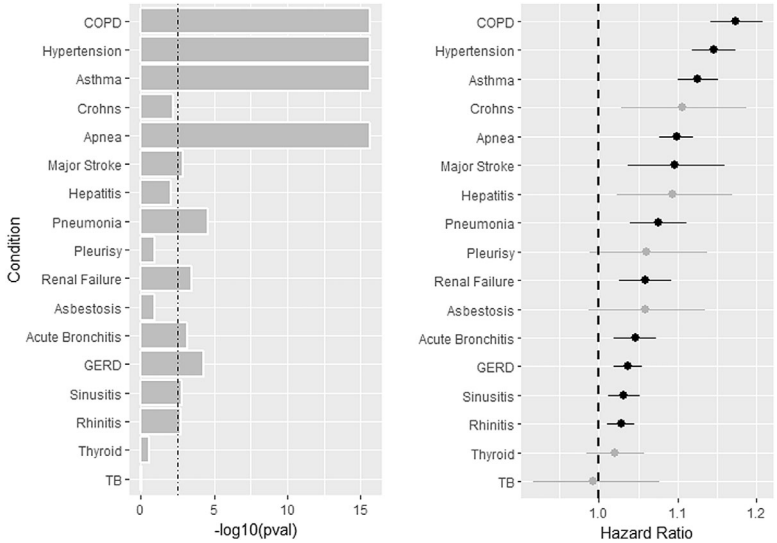

FIGURE 4 Left Panel: Negative log-transformed *P*-values of association between risk score (at Visit 1) and time to disease onset. The dotted line represents the Bonferroni-corrected *P*-value significance threshold for the 17 tests (1 per disease). Significance at the corrected threshold was found for 11 out of 17 conditions. For COPD, hypertension, asthma, and apnea, exact *P*-values could not be obtained since they fell below the numeric limit in the R statistical software, that is, $P < 2.2 \times 10^{-16}$. Right Panel: Forest plot showing hazard ratios (and 95% confidence intervals) for the risk score effects for the 17 conditions; point estimates and CIs are shown in black for the 11 effects found significant after Bonferroni correction (the other six are shown in gray). The dotted line represents a hazard ratio of 1

In these stratified analyses, we excluded the six conditions which were not significantly associated with the risk index in the main (non-stratified) analyses, leaving the following 11 conditions: GERD, hypertension, rhinitis, sinusitis, sleep apnea, asthma, acute bronchitis, renal failure, COPD, pneumonia, and major stroke. Within each stratum of each stratification variable, models were fit for these 11 conditions and adjustment for multiple testing was carried out using the Bonferroni procedure. *P*-values were deemed significant if they fell below a Bonferroni threshold of 0.00455 (=0.05/11). The results are summarized in Tables S4a-S6c in the appendix. Overall, the stratification analyses reveal that, within certain subgroups, not all 11 conditions are associated with the risk index. This could be due to the reduced sample size that results from stratifying the cohort into subgroups. Indeed, we observed that across the strata examined, the number of conditions showing significant association with the risk index was proportional to stratum sample size. Despite this, the stratified analyses showed a fairly consistent pattern of association involving a core set of conditions—hypertension, asthma, COPD, and sleep apnea.

3.4 | Association of risk index with self-assessed physical wellbeing

Quantile regression was used to test the association of the risk score with the PCS SF-12. The analysis was carried out on the subset of the full analytic sample (n = 22 833) who had completed at least two visits by December 31, 2015. For these 15 387 individuals, summary statistics are provided in Table 5. Mean age at visit 2 was 47.1 years. Approximately half (47.5%) enrolled in the program between 2002 and 2005. At visit 2, a majority (54%) reported never having smoked. The distributions of sex, race/ethnicity, pre-9/11 occupation, and exposure severity among these individuals were not substantially different from those observed for the full analytic sample (Table 2).

The mean value of the SF-12 Physical Composite Score at visit 2 was 43.9 (median = 45.8). The histogram in Figure 5 depicts the distribution of this score among the 15 387 GRC members used in the cross-sectional analysis testing the association of the risk index with SF-12 PCS at visit 2. The distribution deviates substantially from normality and appears bimodal. For this reason, quantile regression was used for the analysis, the results of which are summarized in Figure 6.

Figure 6 shows the regression coefficients quantifying the association of the risk score with various quantiles of SF-12 PCS. The SF-12 PCS instrument was designed so that higher scores correspond to better health status. On the other hand, for our risk index, higher values (ie, higher mortality risk) mean worse health. Therefore, in the figure, the significantly negative risk index coefficients (observed across all quantiles of SF-12 PCS) imply strong agreement between our risk index and self-assessed health status.

4 | DISCUSSION

We have developed and validated an index that represents an aggregation of the health information from various physiological parameters—blood biochemistry profiles (from the metabolic panel, lipid panel, and complete blood count), pulmonary function tests and anthropometric measures. The index was developed via a data-driven approach, applying a machine-learning algorithm to data from

TABLE 5 Summary statistics for sample used in Visit 2 analysis of relationship between risk score and self-rated physical health

Variables	Summary statistics
Number of GRC members	15 387
SF-12 physical composite score (mean)	43.9
Age at Visit 2 (mean)	47.1
Sex	
Female	2388 (15.5%)
Male	12 999 (84.5%)
Race/Ethnicity	
White	8785 (57.1%)
Black	2034 (13.2%)
Hispanic	3991 (25.9%)
Other	455 (3%)
Missing	122 (0.8%)
Smoking status at Visit 2	
Never	8312 (54%)
Former	3812 (24.8%)
Current	1596 (10.4%)
Missing	1667 (10.8%)
Occupation at 9/11	
Construction	3510 (22.8%)
Protective	7394 (48.1%)
CM & IRG	1308 (8.5%)
Other	3027 (19.7%)
Missing	148 (1%)
Entry year (Visit 1 year)	
2002-2005	7308 (47.5%)
2006-2008	4783 (31.1%)
>2008	3296 (21.4%)
Exposure severity	
Low	2124 (13.8%)
Intermediate	9522 (61.9%)
High/Very high	3290 (21.4%)
Missing	451 (2.9%)

NHANES III. Our validation analyses showed that the value of this index computed at an individual's first visit to the program is predictive of subsequent mortality and development of multiple comorbidities. The index, therefore, can be utilized as a metric of life expectancy and general morbidity risk. Particularly noteworthy is the observed association between the SF-12 measure of physical wellbeing and our index, indicating that the latter could potentially serve as a measure of overall physiological health. Because the index represents an aggregation of a battery of clinical lab tests that assess several key physiological parameters, it can be seen as a composite measure of physiological functioning. The physiological parameters/biomarkers used in this index are derived from routine clinical laboratory tests,

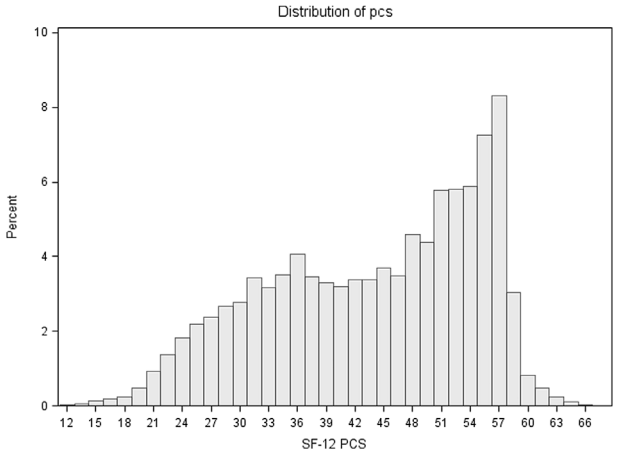

FIGURE 5 Distribution of Physical Composite of SF-12

indicators of an individual's underlying physiological health and disease risk. This index has the potential to be useful for identifying individuals at the early stages of disease, when levels of clinical parameters/biomarkers are only slightly elevated but perhaps still within clinical reference ranges considered normal. As a multi-biomarker composite, the index is by design sensitive to the cumulative effect of mild deviations of multiple health biomarkers from their ideal ranges, and could therefore identify pathological processes in their early stages of development. For GRC members with multiple monitoring visits over the years, the index could be useful as a standardized metric for tracking their health over time. In this regard, it is a potentially valuable tool for studying the dynamics of health trajectories among the cohort. We have discussed the potential usefulness of our index as a clinical tool for health risk monitoring among the GRC. But one important

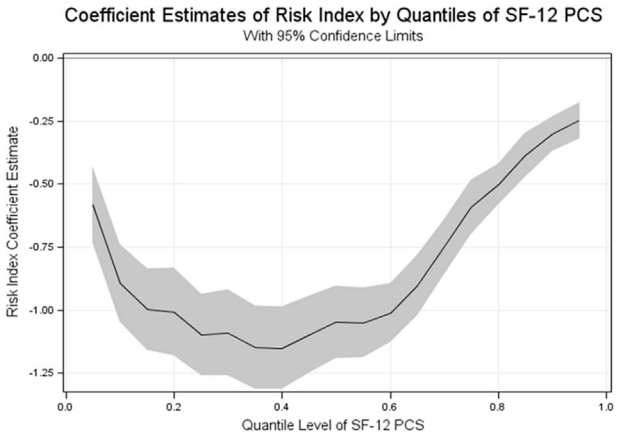

FIGURE 6 Quantile process plot showing the variation in the Risk Index coefficient estimate over different quantiles of the SF-12 PCS. A series of quantile regressions were performed where the outcome variable was a specific quantile of PCS and the independent variables were the risk index and demographic covariates. The coefficient estimates for the risk index across these models are depicted by the black solid line, with confidence bands shown in gray

question is whether the physiological health risk the index is designed to quantify is related to 9/11 exposure. In this study, we found associations between the risk index and a few health conditions which have been designated as 9/11 related, for example, asthma, GERD, and chronic rhinosinusitis. Chronic sinusitis and rhinitis are common diagnoses among cohort members but our risk index showed only a weak, marginally significant association with these conditions. Also, it must be noted that our analyses establishing these associations relied on GRC member self-reports of these conditions, so some of the reported cases may not be directly linked to WTC efforts. Establishing that a case is significantly related to 9/11 exposure requires a formal process of certification. In this study, we did not use data on certified cases. Among the analytic sample, there were a total of 261 confirmed deaths. About a fifth of these were non-disease related (eg, injuries & violence). Roughly 36% were cancer-related, 20% were related to heart disease, 6% were related to respiratory diseases, 5% were infection-related, 4% were diseases related to endocrine/metabolic disorders, and 9% were due to other causes. There is so far no clear and established link between WTC-related exposure and mortality among the general responder cohort.[60]

Computation of the risk index for GRC members is complex but easily automated, and can be applied to available lab/spirometry/physical exam data at monitoring visits. This would yield a raw risk index score for each individual at each visit. While higher scores generally connote greater health risk, the value of the raw index score itself has little intrinsic meaning. But there are a variety of approaches for translating these raw scores into interpretable measures. One approach would be to compare an individual's raw score against the score distribution for the general responder cohort in order to categorize the individual into one of multiple risk groups. In this study, we attempted to do this by using a crude categorization based on tertiles of the risk score in the full analytic sample—stratifying individuals into low, medium, and high risk groups. Our results showed that, compared to the low and medium risk groups, the high risk group had significantly worse survival trends (after age and sex adjustment). But, the difference in survival trends between low and medium risk groups was not significant. However, in practice, risk groups need not be defined by tertiles (or other quantiles); they could instead be derived in such a way as to produce groups with distinct and significantly different survival trends. This is a subject of ongoing research. As an alternative approach, the raw risk score can be compared to age- and sex-specific reference curves to derive a percentile-based score. For each individual, this score would represent the percentage of GRC members of similar age and sex whose raw risk scores fall below that of the individual in question. This has a more intuitive interpretation because it gives a sense of an individual's physiological health status relative to their demographic peers. The practical utility of this risk index as a clinical screening tool for GRC members comes with some caveats. In presenting these risk scores to GRC members, there is a potential for misinterpretation. This is a general concern with all clinical risk prediction tools. Because of their inherently probabilistic nature, presenting them to patients requires careful explanation of what they imply. This may be difficult, particularly for those patients who are not accustomed to thinking in probabilistic terms. It is important to stress to a patient that their score on the risk index does not represent a deterministic claim about their future, rather, it provides merely a probabilistic estimate of mortality and comorbidity risk. Further, presenting the risk score in terms of age- and sex-specific percentiles (as discussed above) allows a straightforward interpretation.

This study has a number of limitations. Our analyses focused on assessing the usefulness of the index when computed at GRC members' first or second monitoring visits. However, the clinical parameters used to compute the index are measured at every visit, so a longitudinal analysis (with multiple values of the index over time) could have been used in this study. We chose instead to use values of the risk index computed at a fixed baseline time point (Visit 1) as predictors of mortality or disease onset. This prospective design utilizes only a portion of the complete laboratory exam data available on GRC members. Further, the timing of Visit 1 among responders in our analytic sample varies quite extensively (from 2002 to 2015). All our analyses included an adjustment for the entry period in order to account for heterogeneity in the condition of responders at their first visits. However, stratification by entry period may be more appropriate in this case. We could not carry this out due to extremely low mortality rates within strata. The primary reason for the prospective design used in this study was to facilitate the assessment of the index in terms of its strength of association with various endpoints. Using longitudinally observed biomarker and health data will require a more complex analytical approach. Furthermore, due to attrition and loss to follow-up, analysis of the longitudinal biomarker data for this cohort (particularly in relation to survival outcomes) will require careful consideration of potential sources of bias. Future studies will be devoted to addressing these challenges. We, therefore, consider the simple, prospective design adopted here as an essential first step to carrying out a more comprehensive evaluation of the prognostic capabilities of the index. A further limitation of our study concerns the high censoring rate for mortality among the WTC cohort: among the 21 106 individuals for whom mortality status could be ascertained, there were only 261 confirmed cases of death (1.2%) as of 12/31/2011 (NDI linkage end date). On 9/11/2001, the average age of individuals in our analytic sample was about 39.3 years, so the low mortality rate as of 2011 is not unexpected. We plan to perform another NDI linkage to reassess mortality status at a more recent time, particularly for newer GRC members who enrolled after 2011. But despite the rather high (98.8%) censoring rate for mortality in this study, our index still demonstrated a strong association with mortality. One caveat related to this and other analyses performed in this study is worth highlighting here. Interpretation of the significant effects reported in this study should take into account the rather large sample sizes the analyses were based on. Another issue is that we tested the association of our risk index with only a fraction of the health conditions observed among responders. The 17 conditions selected for this study were chosen based on availability of sufficient curated data on them. While some of these conditions (particularly GERD, hypertension, and chronic rhinosinusitis) are among the most

commonly diagnosed within the GRC cohort, they represent only a fraction of all reported conditions to date. Future work will focus on examining other relevant conditions. Related to this, one limitation particularly worth highlighting has to do with our study's reliance on self-report for determining the occurrence (and timing) of disease onset in our sample. Valid concerns could be raised with regard to the accuracy of this form of disease ascertainment. Also, potential conflicts could exist between self-reports and clinical information (physical exams, lab tests, etc.) collected on these individuals. This is particularly important at Visit 1, due to the prospective design of our analyses of the conditions. Recall that, using clinical parameters measured at Visit 1 (baseline), we compute the risk index and test its association with future disease onset. It is therefore critical to screen out those who already have the condition at baseline. This was carried out for all conditions via self-report, but for hypertension, we also utilized blood pressure measurements from the physical examination performed at Visit 1. This additional screening criterion was applied for hypertension since it is a largely asymptomatic condition that tends to be underdiagnosed. As discussed in the section 2, we excluded individuals whose blood pressure fell within the hypertensive range at Visit 1 (even if they did not report having been diagnosed with it). As for subsequent visits, conflicts between self-report and clinical diagnostic measures at these visits were not accounted for in our analyses, which is a limitation of our study. It must, however, be noted that results from routine physicals and lab tests cannot replace a proper physician diagnosis involving thorough examination. To the extent that they are accurate, lab test results suggest, but do not confirm, the presence of a condition. Thus, for scenarios where diagnoses are self-reported (such as this study), it is debatable whether lab test/physical exam results should always take precedence in cases of conflict. One other consideration is the fact that certain conditions may be misdiagnosed, which is another factor influencing the reliability of self-report. In particular, two or more conditions with similar clinical features and symptoms could be diagnostically conflated. For example, COPD and asthma have overlapping sets of symptoms and may sometimes be conflated.[66] We performed a sensitivity analysis wherein COPD and asthma were combined into a composite condition and the association of the risk index with this combination was assessed using a Cox Proportional Hazards Model. We adjusted for demographic covariates, smoking status, entry year, and exposure severity. The results indicated that the risk index is significantly associated with the COPD/asthma composite (Hazards Ratio: 1.14 [95%CI: (1.11, 1.16)]).

The transition from healthy to disease states involves a gradual divergence of physiological parameters away from their "normal" (homeostatic) levels; a process described by the generic term "physiological dysregulation."[41] Measuring the extent of this dysregulation in an individual can allow assessment of the cumulative burden of current diseases and, for those in earlier stages of disease development, can also predict future disease risk. In this study, we have attempted to do this for the WTC general responder cohort by developing a mortality-calibrated index that represents a composite of various physiological markers of health. Since increasing dysregulation in physiological systems is accompanied by increasing mortality risk, our index serves as an excellent proxy of cumulative physiological dysregulation and morbidity risk.

AUTHORS' CONTRIBUTIONS

GAB conceptualized and designed the study, and performed all statistical analyses. MAC, DJH, BJL, JMM and IGU all contributed to data acquisition. MS, JRK, CRD contributed to data cleaning and processing. SLT, ACT, RGL contributed to interpretation of the results and modification of study design. GAB drafted the first version of the manuscript and all other authors contributed to substantial revision of the manuscript, and approved the final version of the paper to be published.

ACKNOWLEDGMENTS

We are grateful to: the responders; all past and present staff of the Programmes; the work force, community and volunteer organization programme stakeholders; and the philanthropic sources that funded treatment from 2003 to 2006. We would also like to give special thanks to Dr. Tania Carreón-Valencia and Dr. Geoffrey M. Calvert for their helpful reviews.

FUNDING

This work was supported via contracts and grants 200-2002-00384, U10-OH008216/23/25/32/39/75 and 200-2011-39356/61/77/84/85/88 from the Centers for Disease Control and Prevention/National Institute for Occupational Safety and Health (CDC/NIOSH). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of CDC/NIOSH.

ETHICS APPROVAL AND INFORMED CONSENT

The World Trade Center Health Programs were initially approved by the Institutional Review Board of The Mount Sinai School of Medicine, and subsequently by both the IRB of the Icahn School of Medicine at Mount Sinai and IRBs of the clinical sites listed in the author affiliations. The Health Programs obtained the signed consent of all participants. All procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

DISCLOSURE (AUTHORS)

Dr. Jacqueline Moline has served as an expert witness in cases involving asbestos tort litigation; Dr. Moline's institution has received funds from the Transport Workers Union and the United Federation of Teachers (for consultancy), Medscape (for development of educational presentations on WTC health effects), and has current/pending grants from NIOSH (as part of the WTC Health Program) and FASNY

(Firemen's Association of the State of New York). Other authors declare no conflicts of interest.

DISCLOSURE BY AJIM EDITOR OF RECORD

Steven B. Markowitz declares that he has no conflict of interest in the review and publication decision regarding this article.

DISCLAIMER

None.

ORCID

Ghalib A. Bello http://orcid.org/0000-0002-7893-743X
Roberto G. Lucchini http://orcid.org/0000-0002-9723-0237

REFERENCES

1. Webber M, Moir W, Crowson C, et al. Post-September 11, 2001, incidence of systemic autoimmune diseases in world trade center-exposed firefighters and emergency medical service workers. *Mayo Clin Proc*. 2016;91:23–32.
2. Yip J, Webber M, Zeig-Owens R, et al. FDNY and 9/11: clinical services and health outcomes in World Trade Center-exposed firefighters and EMS workers from 2001 to 2016. *Am J Ind Med*. 2016;59:695–708.
3. Dasaro C, Holden W, Berman K, et al. Cohort profile: world trade center health program general responder cohort. *Int J Epidemiol*. 2015;46:1–8.
4. Gargano L, Nguyen A, DiGrande L, Brackbill R. Mental health status of World Trade Center tower survivors compared to other survivors a decade after the September 11, 2001 terrorist attacks. *Am J Ind Med*. 2016;59:742–751.
5. Li J, Cone J, Alt A, et al. Performance of self-report to establish cancer diagnoses in disaster responders and survivors, world trade center health registry, New York, 2001–2007. *Public Health Rep*. 2016;131:420–429.
6. Wilkenfield M, Fazzari M, Segelnick J, Stecker M. Neuropathic symptoms in world trade center disaster survivors and responders. *J Occup Environ Med*. 2016;58:83–86.
7. Crowley L, Herbert R, Moline J, et al. Sarcoid like granulomatous pulmonary disease in world trade center disaster responders. *Am J Ind Med*. 2011;54:175–184.
8. Boffetta P, Zeig-Owens R, Wallenstein S, et al. Cancer in world trade center responders: findings from multiple cohorts and options for future study. *Am J Ind Med*. 2016;59:96–105.
9. Clouston S, Kotov R, Pietrzak R, et al. Cognitive impairment among World Trade Center responders: long-term implications of re-experiencing the 9/11 terrorist attacks. *Alzheimers Dement (Amst)*. 2016;4:67–75.
10. Hashim D, Boffetta P, Galsky M, et al. Prostate cancer characteristics in the world trade center cohort, 2002–2013. *Eur J Cancer Prev*. 2016; [Epub ahead of print].
11. Martin C. 2016. *Multimorbidity: Through a Glass Darkly*. Switzerland: Springer.
12. McPhail S. Multimorbidity in chronic disease: impact on health care resources and costs. *Risk Manag Healthcare Policy*. 2016;9: 143–156.
13. Nunes B, Flores T, Mielke G, Thume E, Facchini L. Multimorbidity and mortality in older adults: a systematic review and meta-analysis. *Arch Gerontol Geriatr*. 2016;67:130–138.
14. Wisnivesky J, Teitelbaum S, Todd A, et al. Persistence of multiple illnesses in World Trade Center rescue and recovery workers: a cohort study. *Lancet*. 2011;378:888–897.
15. Britt H, Harrison C, Miller G, Knox S. Prevalence and patterns of multimorbidity in Australia. *Med J Aust*. 2008;189:72–77.
16. Barnett K, Mercer S, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multi-morbidity and implications for health care, research and medical education: a cross-sectional study. *Lancet*. 2012;380:37–43.
17. Faner R, Cruz T, Lopez-Giraldo A, Agusti A. Network medicine, multimorbidity and the lung in the elderly. *Eur Repir J*. 2014;44:775–788.
18. van der Heijden M, Schroen Y. Global advances in health and medicine through systems biology: an example from the Netherlands. *Global Adv Health Med*. 2012;1:92–93.
19. Federoff H, Gostin L. Evolving from reductionism to holism: is there a future for systems medicine? *JAMA*. 2009;302:994–996.
20. McEwen BS. Stress, adaptation, and disease. Allostasis and allostatic load. *Ann NY Acad Sci*. 1998;840:33–44.
21. Seeman T, McEwen B, Rowe J, Singer B. Allostatic load as a marker of cumulative biological risk: macArthur studies of successful aging. *PNAS*. 2001;98:4770–4775.
22. Karlamangla A, Singer B, McEwen B, Rowe J, Seeman T. Allostatic load as a predictor of functional decline: macArthur studies of successful aging. *J Clin Epidemiol*. 2002;55:696–710.
23. Singer B, et al. 2004. *Operationalizing Allostatic Load. Allostasis, Homeostasis, and the Costs of Physiological Adaptation. J. Schuli*. Cambridge, UK: Cambridge University Press. pp. 113–149.
24. Gruenewald T, Seeman T, Karlamangla A, Sarkisian C. Allostatic load and frailty in older adults. *J Am Geriatr Soc*. 2009;57:1525–1531.
25. Juster R-P, McEwen B, Lupien S. Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neurosci Biobehav Rev*. 2010;35:2–16.
26. Booth T, Royle N, Corley J, et al. Association of allostatic load with brain structure and cognitive ability in later life. *Neurobiol Aging*. 2015;36:1390–1399.
27. Buckwalter J, Castellani B, McEwen B, et al. Allostatic load as a complex clinical construct: a case-based computational modeling approach. *Complexity*. 2015;21:291–236.
28. Nobel L, Roblin D, Becker E, et al. Index of cardiometabolic health: a new method of measuring allostatic load using electronic health records. *Biomarkers*. 2016;22:394–402.
29. Nakamura E, Miyao K, Ozeki T. Assessment of biological age by principal component analysis. *Mech Ageing Dev*. 1988;46:1–18.
30. Klemera P, Doubal S. A new approach to the concept and computation of biological age. *Mech Ageing Dev*. 2006;127:240–248.
31. Levine M. Modeling the rate of senescence: can estimated biological age predict mortality more accurately than chronological age? *J Gerontol A Biol Sci Med Sci*. 2013;68:667–674.
32. Kim S, Jazwinski S. Quantitative measures of healthy aging and biological age. *Healthy Aging Res*. 2015;4:pii: 26.
33. Putin E, Mamoshina P, Aliper A, et al. Deep biomarkers of human aging: application of deep neural networks to biomarker development. *Aging*. 2016;8:1021–1033.
34. Cohen A, Milot E, Morissette-Thomas V. Statistical distance: A promising measure of physiological dysregulation in biodemographic studies. XXVII International Union for the Scientific Study of Population, Busan, South Korea. 2013.
35. Cohen A, Milot E, Yong J, et al. A novel statistical approach shows evidence for multi-system physiological dysregulation during aging. *Mech Ageing Dev*. 2013;134:110–117.
36. Cohen A, Milot E, Li Q, et al. Cross-population validation of statistical distance as a measure of physiological dysregulation during aging. *Exp Gerontol*. 2014;57:203–210.

37. Milot E, Morissette-Thomas V, Li Q, et al. Trajectories of physiological dysregulation predicts mortality and health outcomes in a consistent manner across three populations. *Mech Ageing Dev*. 2014;141-142: 56–63.
38. Cohen A. Physiological and comparative evidence fails to confirm an adaptive role for aging in evolution. *Curr Aging Sci*. 2015;8: 14–23.
39. Cohen A, Li Q, Milot E, et al. Statistical distance as a measure of physiological dysregulation is largely robust to variation in its biomarker composition. *PLoS ONE*. 2015;10:e0122541.
40. Arbeev K, Cohen A, Arbeeva L, et al. Optimal versus realized trajectories of physiological dysregulation in aging and their relation to sex-Specific mortality risk. *Front Public Health*. 2016;4:3.
41. Cohen A. Complex systems dynamics in aging: new evidence, continuing questions. *Biogerontology*. 2016;17:205–220.
42. Bello G, Dumancas G. A random forest-Induced distance-Based measure of physiological dysregulation. *Curr Aging Sci*. 2017;10: 270–281.
43. Seeman T, Singer B, Rowe J, Horwitz R, McEwen B. Price of Adaptation—Allostatic load and its health consequences. *Arch Intern Med*. 1997;157:2259–2268.
44. Levine M, Crimmins E. A comparison of methods for assessing mortality risk. *Am J Hum Biol*. 2014;26:768–776.
45. Horne B, May H, Muhlestein J, et al. Exceptional mortality prediction by risk scores from common laboratory tests. *Am J Med*. 2009;122: 550–558.
46. Bello G, Dumancas G, Gennings C. Development and validation of a clinical risk-assessment tool predictive of all-cause mortality. *Bioinform Biol Insights*. 2015;9:1–10.
47. Horne B, May H, Kfoury A, et al. The Intermountain Risk Score (including the red cell distribution width) predicts heart failure and other morbidity endpoints. *Eur J Heart Fail*. 2010;12:1203–1213.
48. Austin P, van Walraven C. The Mortality Risk Score and the ADG Score: two points-based scoring systems for the Johns Hopkins Aggregated Diagnosis Groups (ADGs) to predict mortality in a general adult population cohort in Ontario. *Canada. Med Care*. 2011;49: 940–947.
49. May H, Anderson J, Muhlestein J, et al. Addition of albumin, bilirubin, WBC differential count, and other routinely collected laboratory tests to the Intermountain Risk Score improves the predictive ability of death among a general healthcare population. *J Am Coll Cardiol*. 2012;59.
50. Horne B, Anderson J, Muhlestein J, Ridker P, Paynter N. Complete blood count risk score and its components, including RDW, are associated with mortality in the JUPITER trial. *Eur J Prev Cardiol*. 2015;22:519–526.
51. May H, Reiss-Brennan B, Brunisholz KD, Horne B. Addition of depression severity indices to the intermountain risk score for predicting 3-Year risk of chronic disease: the mental health integration risk score. *Circulation*. 2016;134.
52. Plan and operation of the Third National Health and Nutrition Examination survey, (1994). 1988–94: seriesC 1: programs and collection procedures. *Vital Health Stat*. 1:1–407.
53. National Center for Health Statistics (2013). National Death Index User Guide. from http://www.cdc.gov/nchs/data/ndi/NDI_Users_Guide.pdf
54. Friedman J. Greedy function approximation: a gradient boosting machine. *Ann Statist*. 2001;29:1189–1232.
55. Elith J, Leathwick J, Hastie T. A working guide to boosted regression trees. *J Animal Ecol*. 2008;77.
56. McPherson R, Pincus M. 2011. *Disease/organ panels*. Philadelphia, PA: Elsevier Saunders.
57. Vajpayee N, et al. 2011. *Basic examination of blood and bone marrow*. Philadelphia, PA: Elsevier Saunders.
58. National Center for Health Statistics (2013). NCHS 2011 Linked Mortality Files Matching Methodology. from https://www.cdc.gov/nchs/data/datalinkage/2011_linked_mortality_file_matching_methodology.pdf.
59. Ridgeway G. 2007; Generalized boosted models: A guide to the gbm package. from http://132.180.15.2/math/statlib/R/CRAN/doc/vignettes/gbm/gbm.pdf
60. Stein C, Wallenstein S, Shapiro M, et al. Mortality among world trade center rescue and recovery workers, 2002–2011. *Am J Ind Med*. 2016;59:87–95.
61. Gandek B, Ware J, Aaronson N, et al. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. *J Clin Epidemiol*. 1998;51:1171–1178.
62. Busija L, Pausenberger E, Haines T, et al. Adult measures of general health and health-related quality of life. *Arthritis Care Res*. 2011;63: S383–S412.
63. Koenker R. 2005. *Quantile Regression*. Cambridge, UK: Cambridge University Press.
64. R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
65. Gail M, Byar D. Variance calculations for direct adjusted survival curves, with applications to testing for no treatment effect. *Biometric J*. 1986;28:587–599.
66. Bellia V, Battaglia S, Catalano F, et al. Aging and disability affect misdiagnosis of COPD in elderly asthmatics: the SARA study. *CHEST*. 2003;123:1066–1072.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Bello GA, Teitelbaum SL, Lucchini RG, et al. Assessment of cumulative health risk in the World Trade Center general responder cohort. *Am J Ind Med*. 2018;61:63–76. https://doi.org/10.1002/ajim.22786