Supplemental File S1

Details of Statistical Modeling Plan

The hypotheses underlying our framework are that: (A) dynamic measures are key indicators of resilience capacity, (B) resilience capacity is a primary driver of response to stressors, and (C) the dynamic functioning of key physiologic systems, in addition to pre-stressor biologic, health, and psychosocial characteristics, determines resilience capacity. In keeping with our study aims developed to test these hypotheses and build measures by which to assess resilience capacity in advance of stressors, we have identified four specific statistical modeling goals that are detailed here. Analytical approach for each of these goals is described here.

(1) To measure physiologic resilience capacity (PRC) using the *baseline* data from static and dynamic assessments. Our conceptual framework hypothesizes physical resilience capacity (oval in Figure 1) as a physiological system property that is not directly quantifiable, but can be implicitly described through a combination of dynamic stimulus test metrics and static surrogate measures. To our knowledge, our study is the first to formally address this hypothesis (C above) or to collect multiple stimulus response measures in the same individual, hence this task is both central to our study and among its highest innovations. According to our theory as to the dynamic nature of multisystem interactions involved in physical resilience, the connection between physiologic capacity and the dynamic measures ought to be rigorously described by differential equations. Ideally, then, models used to analyze dynamic stimulus test data should build in such equations. This, however, will not be broadly possible for us due to the limited frequency of repeated measurement of system response that is feasible in clinical settings. Therefore, we will implement three other approaches to combine the data collected into a categorization or score that can then be used to evaluate PRC association with, and ability to predict, resilience phenotypes and clinical outcomes.

First, we will derive a: “PRC-deficit” by developing a simple sum (index) of deficits. For the PRC assessments with established normal ranges, we will score those outside the normal range as 1 and those within the normal range as 0. For continuous measures without established normal range we will score those in the “worst,” say, tertile as 1 and others as 0. Second, we will generate a “PRC-scale” which implements a simplified version of our theoretical framework, using latent variable models. These models will hypothesize summary physiologic states or processes (e.g. physiologic “steady state” and “adaptation”) and consider the test results as indicators of these [31–34]. Estimates of the latent variables in this model then provide estimates of the physiologic states or processes. Third, a machine learning (ML)-based “PRC-ML” will be generated. The first two approaches articulated in this section ignore the ultimate goal, which is to identify individuals whose ultimate stressor responses are non-resilient. In this third approach, we will create a resilience capacity measure from a combination of dynamic stimulus and static surrogate measures that best predict resilience phenotypes. ML methods will be implemented to develop, cross-validate, and compare candidate predictive algorithms (as an example for technical readers, using Gaussian lasso regression). PRC thus estimated will be a risk score or classifier, a combination of the PRC measures collected by which to forecast phenotypes or outcomes. PRC-deficit and PRC-scale (but not PRC-ML) will be employed to evaluate hypotheses of association with resilience phenotypes and outcomes; because PRC-ML will target phenotypes in its creation, we will evaluate it only for the predictive purpose of identifying persons at risk for adverse clinical outcomes.

(2) To characterize resilience phenotypes of stressor-response: We consider multiple assessments identified as promising resilience phenotype candidates (Table 2). Prioritization of these for further development will focus on their feasibility, reliability, sensitivity to a stressor,

between-subject heterogeneity in stress response, and target of measurement with respect to systemic vs. organ-specific functioning. To assess feasibility, we will audit variable- and instrument-level missing data, distinguishing missingness due to health vs. other reasons. Both a large proportion of missing data and disproportional missingness in selective subset(s) of the study sample would evidence poor feasibility and/or generalizability. To assess reliability, for continuous measures, mixed effects models [35] will be used with random intercepts and random coefficients for individual-specific time trend (e.g., linear time slope); and reliability can then be calculated as ratio of the variance of the random intercept over the sum of the variance of the random intercept and the residual variance estimated from the mixed effects models [35]. To assess sensitivity to a stressor, stressor-induced change in a phenotypic indicator (e.g., SPPB) before and after the stressor will be analyzed using regression models that account for stressor magnitude and baseline demographic characteristics.

Next, we will develop phenotypes of stressor response by characterizing trajectories of phenotypic measures before-to-after the stressor, allowing heterogeneity between individuals and for curvilinear shape (notably, an initial decline followed by a rebound). Mixed effects models and functional principal components analysis adapted for sparsely repeated data (fPCA) [36,37] are two methods we will employ to achieve this. Each prioritized phenotypic measure will be analyzed separately first. Then, the inter-dependence between the different phenotypic measures will be analyzed using multivariate versions of REM [38] and fPCA [39] to model joint patterns of trajectories across phenotypic measures. Versions of the former (multivariate growth mixture models; [40,41] will be used to identify groups of individuals with similar patterns of trajectories across measures.

(3) To characterize the association and predictive accuracy/precision of physical resilience capacity measures for resilience phenotypes and outcomes following the stressor. Development in (1) will yield at least three candidates for comparison—PRC-deficit, PRC-scale, and PRC ML. We will assess the degree to which PRC-deficit and PRC-scale independently predicts the phenotypic measures developed in (2), after adjusting for determinants already in common clinical use, including age, sex, multimorbidity and BMI. We also will assess the degree to which PRC may moderate the stress-response by studying the interaction between resilience capacity and stressor magnitude, and evaluate which of the PRC estimates most accurately predicts resilience phenotypes and longer-term outcomes—using cross-validation to estimate performance in future application. As a final evaluation, we will build and compare a predictive algorithm incorporating multiple metrics spanning dynamic testing and static measures using machine learning techniques. Such an approach recognizes that no one summary may even approximately optimize prediction. Estimates of the sensitivity and specificity for identifying persons at risk for adverse outcomes, and precision for forecasting the resilience phenotype, will be produced.

4) To explore age-related biological mechanisms potentially contributing to physiologic resilience capacity. The SPRING investigative team is measuring multiple variables from several potentially important biological domains known to change with aging. These include molecular (senescence cell surface markers, metabolomic measures, epigenetic markers), physiological (baseline inflammatory cytokines, urine catecholamines, ghrelin and other hormones) and clinical laboratory measurements (complete blood count, metabolic panel). Some of these measures have long standing predictive validity for resilience phenotypes and outcomes (i.e., inflammatory cytokines, hemoglobin) and others are new measures that have not been extensively studied in older adults (T-cell phenotypes, senescent markers). Given that the dynamic systems measures are likely influenced by age-related biological alterations, documenting associations of biological variables with dynamic physiological variables will have value. Given the early stage of discovery, analyses will appropriately account for multiplicity, for example, using multiple comparisons corrections and penalized regression approaches. These analyses will help generate – hypothesis-driven analyses in future studies.

Cross-cutting Methodological Challenges

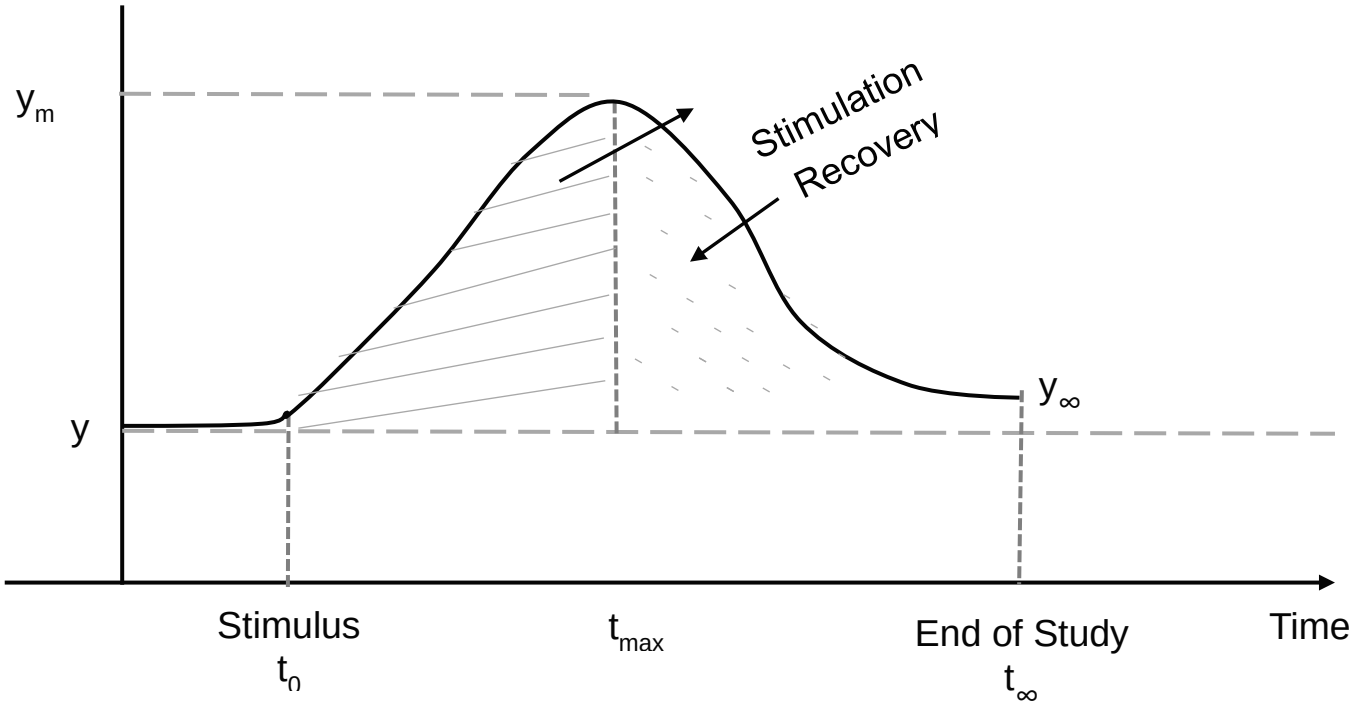
There are several methodological challenges that cut across these analytical modeling goals. Firstly, the prevalence of informatively missing data and censoring of trajectories due to study dropout in each clinical stressor, i.e., patients who are sicker tend to miss entire visits or a subset of study measures. We will prioritize estimators that are robust to data missing at random (e.g., maximum likelihood estimator in the case of REM) and conduct analyses employing informative imputation when warranted (e.g., assigning the lowest level of function for inability to perform handgrip test due to arthritis pain), multiple imputation, and sensitivity analyses. Then there is the difficulty of handling death in statistical models. Death is clearly a non-resilient outcome, but it poses a challenge for modeling trajectories of resilience phenotypes. Joint trajectory / survival models are an option here. Finally, it would be useful to examine whether we can identify commonalities across the 3 clinical stressors; for example, common predictors of resilience capacity, and common resilience phenotypes. We can gain power if the data can be pooled across clinical stressors. To this end, a methodological challenge is to evaluate whether the data can be aggregated. It is likely that participants in our study tend to be more robust than non-participants: Sensitivity analyses will be needed to document potential ramifications.

Modeling the Dynamics of Stimulus-Response Experiments

Loss of resilience in homeostatic regulatory systems, which underlies vulnerability to stressors, is fundamentally a dynamic construct. The literature on frailty and resilience has hypothesized that the deleterious changes in the regulatory systems involved in the maintenance of homeostasis may well be subtle and undetectable in the absence of external stressors such as infection, injury, or organ-system-based illness. Consequently, the frail and non-frail would differ more in terms of the dynamics of physiological systems in response to stimuli than they would in terms their baseline status (Buchner and Wagner, 1992; Lipsitz, 2002). Therefore ‘‘resilience’’ is a feature most observable in situations where an external stimulus induces measurable changes in the physiological system under study. Studying a biological system only under basal conditions by measuring static biomarkers cannot address the dynamic properties of that system, i.e., how the system would respond to a challenge, nor does it acknowledge inter-person heterogeneity in basal levels independent of their functional status. Stimulus-response experimentation is a powerful tool to improve our understanding of the vulnerability associated with frailty. Varadhan (2008) proposed a dynamical systems modeling approach, based on the stimulus-response experimental paradigm, to study loss of resilience associated with frailty. This approach was employed in the Women’s Health and Aging Study to demonstrate the difference between older frail and non-frail women in terms of their response to various physiological stimuli (Kalyani 2012; Fried 2021).

The following figure illustrates the features of data from a stimulus-response experiment. The stimulus is applied at a specific time point, t_0. The response of the physiological system is measured in terms of the level of a biomarker, y, as a function of time for an appropriately long duration, t_∞. Maximum response, y_{max}, is observed at time t_{max}, after which the stimulus is resolved and the level of the biomarker decreases and approaches baseline level.

Supplemental Figure 1:

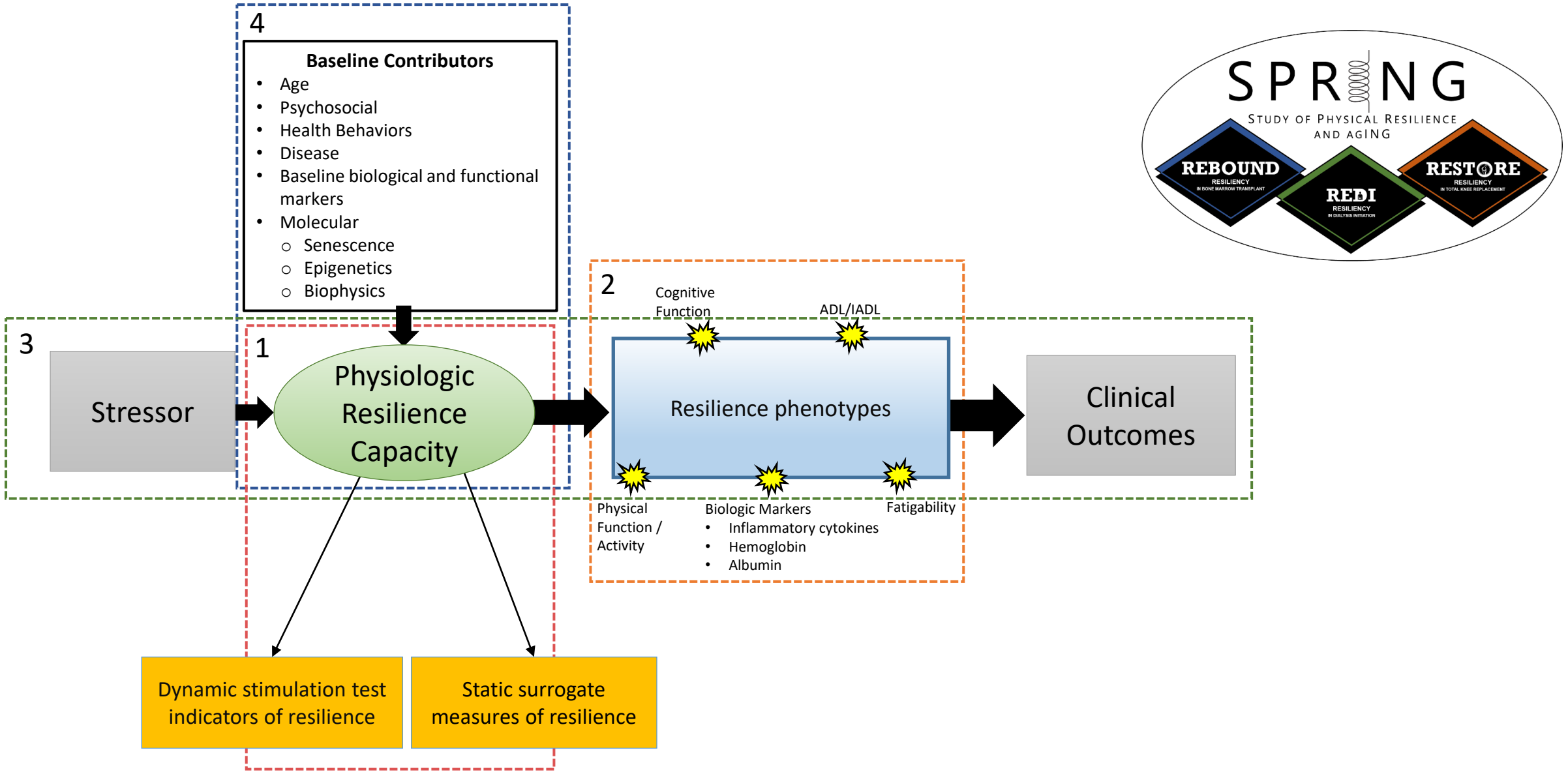


Data, y(t), from the stimulus-response experiment can be analyzed and summarized in a variety of ways. One approach is to use a mathematical model to capture the dynamics of the response. The parameters of the model can provide insight into the underlying physiology. An example of this is the Ackerman model for the oral glucose tolerance test (OGTT) (Ackerman 1962). The Ackerman model is written as: $y(t) = y_0 + A\, e^{-kt}\, \mathrm{Sin}(\omega t)$, where y_0 is the fasting glucose level, A is a constant that represents the magnitude of the stimulation, k is a constant representing the rate of recovery, and ω is the angular frequency of the response. The parameters of the Ackerman model $\theta = \{y_0, A, k, w\}$ can be estimated from discretely sampled data from the OGTT experiment for each individual using the non-linear least squares method. The parameters θ represent the dynamics of the stimulus-response experiment for a given individual, and can be used as potential indicators of the resilience of the individual to actual physical stressors. Varadhan et al. (2008) provides a more in-depth discussion of model-based approaches.

When a physiologically-based mathematical model is not available to model the data, we can use simpler summary measures. These include the baseline level (y_0), recovery level (y_∞), ratio of recovery to baseline (y_0/y_∞), and area-under-the-curve (AUC) of the response. Let $\{t_k, y_k\}$, k=1, 2, …, K, be the glucose levels of an individual collected at K different times. The AUC can be calculated using the trapezoidal rule for integration: $y_{AUC} = 0.5 \sum_{k=2}^{K} (t_k - t_{k-1)}\, (y_k - y_{k-1})$. Kalyani et al. (2012) showed that glucose-AUC was associated with frailty in oldest-old women.

A third approach is to use nonparametric representations y(t), such as functional principal components analysis (fPCA) (Rice and Silverman 1991). Typically, two or three principal components are adequate to capture the variation in the response. We can then use the principal component scores of each individual as summaries of the response.

Supplemental Figure : Annotated Conceptual Framework for Physical Resilience



Supplemental Figure 2 Legend:

Analytic Aims

(1) To measure physiologic resilience capacity using the baseline data from static and dynamic assessments (Red outline)
(2) To characterize resilience phenotypes of stressor-response (Orange outline)
(3) To characterize the association and predictive accuracy/precision of physical resilience capacity measures for resilience phenotypes and outcomes following the stressor (Dark Green outline)
(4) To explore age-related biological mechanisms potentially contributing to physiologic resilience capacity (Dark Blue outline)