Methodology for the 2019 Release

National Health Security Preparedness Index

The National Health Security Preparedness Index tracks the nation's progress in preparing for, responding to, and recovering from disasters and other large-scale emergencies that pose risks to health and well-being in the United States. Because health security is a responsibility shared by many different stakeholders in government and society, the Index combines measures from multiple sources and perspectives to offer a broad view of the health protections in place for the nation as a whole, each state, and the District of Columbia (DC). This document summarizes the measures and computational methodologies used in constructing the 2019 release of the Index.

I. RATIONALE AND HISTORY OF THE INDEX

The U.S. Centers for Disease Control and Prevention (CDC) initiated development of the National Health Security Preparedness Index in 2012 to create a platform for measuring the nation's progress in preparing for, responding to, and recovering from disasters and other large-scale emergencies that pose risks to health and well-being in the United States. As a measurement tool, the Index is designed to summarize levels of preparedness achieved within individual states and for the nation as a whole, with the goal of disseminating and using this information for multiple purposes: (1) to enhance public awareness and understanding of national preparedness components and capabilities; (2) to encourage coordination and collaboration among the multiple sectors and stakeholders that contribute to preparedness capabilities; (3) to inform planning, policy development, and quality improvement activities across the preparedness field; and (4) to stimulate and guide future research on how to measure and improve preparedness and health security.

Supported by CDC, the Index was developed through a broad collaboration of stakeholders led by the Association of State and Territorial Health Officials (ASTHO), the Oak Ridge Associated Universities (ORAU), the University of Pittsburgh Medical Center's Center for Biosecurity, and Johns Hopkins University's Center for Public Health Preparedness. More than 30 additional organizations have contributed to development of the Index over time, including federal, state, and local public health agencies; emergency management agencies; health care organizations; research institutions; and professional associations. Developed as an annual measurement tool, the first edition of the Index was released in December 2013, and a second edition was released in December 2014. This second edition of the tool included a total of 197 measures drawn from more than 60 data sources. The measures were aggregated into domain and subdomain composite measures, and further aggregated into an overall health security score, based on conceptual framework of health security and preparedness developed for the Index.

In January 2015, responsibility for publishing and maintaining the Index transitioned from CDC to the Robert Wood Johnson Foundation (RWJF). RWJF selected a program office based at the University of Kentucky's Center for Public Health Systems and Services Research and its Center for Business and Economic Research to lead efforts to refine and update future editions of the Index. RWJF also appointed a 14-member National Advisory Committee for the Index to provide overall scientific and strategic guidance regarding the Index design, operation, and use. Additionally, the program office established three workgroups to provide operational advice on future updates and revisions to the Index, drawing from a workgroup structure used during development of earlier editions of the Index.

II. REFINEMENTS TO THE INDEX METHODOLOGY AND MEASURES

Upon release of the second edition of the Index in December 2014, the Index Program Management Office initiated a series of activities to examine the existing measurement properties of the Index and to identify strategies for improving the Index as a measurement tool. The aims of these activities were threefold: (1) to determine the construct validity and reliability of the Index domains and subdomains in order to identify strategies for improving these measurement properties; (2) to determine the accuracy of comparisons made across Index domains and subdomains and across states in order to identify strategies for improving the accuracy of comparisons; and (3) to determine valid and feasible methods for supporting longitudinal comparisons of Index values so that changes in health security over time can be accurately tracked.

Four sets of activities were completed by the program office in pursuit of these aims, specifically to:

1. Conduct measurement validity and reliability analyses that examine the performance of existing Index measures in characterizing core preparedness constructs reflected in Index's conceptual framework, including the Index domains and subdomains. These analyses include (a) internal consistency reliability tests performed at the subdomain, domain, and overall Index level; and (b) multi-trait scale analysis tests performed at the subdomain and domain levels;[1]
2. Conduct sensitivity analyses that examine the relative influence of each measure on overall Index results, including the impact of the Index's methods for scaling, imputing, and aggregating individual measures into subdomains, domains, and overall Index values;
3. Solicit ideas for new and modified measures to include in the Index through an Open Call for Measures and through monthly Index Workgroup meetings with content experts and stakeholders in the preparedness field; and
4. Assess the availability, completeness, quality, timeliness, and longitudinal consistency of data sources for existing and proposed new Index measures, including whether data sources are updated at least every three years.[2]

The results of these activities were discussed and refined with preparedness experts and stakeholders during monthly Index Workgroup meetings and during quarterly meetings of the National Advisory Committee. Based on this feedback, the program office drafted a set of recommended changes to the Index measures and methodology that were proposed for implementation as part of the third release of the Index. Proposed changes were developed with the following broad goals for methodological refinement in mind:

- Improve the methods used for grouping and weighting individual measures within domains and subdomains to improve the internal consistency and discriminant power of the Index;
- Consolidate and simplify the overall Index set of measures by reducing unreliable and noisy measures with high levels of measurement error;

---

[1] Staiger D, Dimick JB, Baser O, Fan Z and Birkmeyer JD. Empirically derived composite measures of surgical performance. Medical Care 2009;47: 226-233. Hays RD, Hayashi T. Beyond internal consistency reliability: rationale and user's guide for multitrait analysis program on the microcomputer. Behavioral Research Methods 1990;22(2):167-75.

[2] This criterion for data source periodicity and timeliness is based on the National Quality Forum's measure selection criteria.

- Expand the breadth and composition of the Index by adding new measures reflecting important dimensions of health security and resiliency not currently represented in the Index, including measures that align with established national frameworks for health security, such as the National Health Security Strategy and the preparedness objectives of Healthy People 2020;
- Improve the methods used for scaling individual measures to more accurately reflect the distributional properties of the measures and to enable more accurate comparisons across states and over time;
- Improve the accuracy of the methods used for imputing missing values for Index measures; and
- Incorporate new data and analytic methods that allow for accurate comparisons of Index values over time (trending).

Results of validity and reliability tests were combined with findings from the data source assessments to develop a detailed recommendation about the status of each of the 197 individual measures included in the 2014 release of the Index. For each of these measures, we recommended one of several possible actions: (a) retain the measure as specified on the 2014 Index; (b) modify the way the measure was specified and calculated to improve its validity and/or reliability; (c) reclassify the measure into a different domain and/or subdomain to improve the validity and reliability of the underlying domain and/or subdomain composite measure; or (d) exclude the measure from the next edition of the Index. Measures were recommended for exclusion only if they failed multiple tests of measurement value, including: (i) the measure performed poorly on construct validity and reliability tests at both the domain and subdomain level, as indicated by an adjusted multi-trait item-to-scale correlation coefficient of less than 0.3;[3] (ii) the measure's construct validity and reliability did not improve when reclassified into another domain or subdomain scale; (iii) the measure's validity and reliability had not been established through previously published studies; and (iv) the measure was constructed from a data source that was not updated within a three-year period. By design, these criteria for measure selection and retention place priority on measures that help the Index discriminate health security levels across different domains and subdomains, across U.S. states, and over time.

A document summarizing all proposed changes to the Index methodology and measures was posted for public comment in August 2015. Over a 30-day period, more than 70 comments were received. During October and November 2015, the Index National Advisory Committee reviewed the public comments and made a final set of recommendations regarding changes to the Index methodology and measures for the 2016 release, which were subsequently implemented.

There were fewer proposed changes for the 2017 Index release, but nine new measures were added and four others were removed, resulting in an increase in total measures from 134 to 139. We also added DC as an additional geographic area for all measures and all four years of data. A detailed description of the methodological changes for the 2017 Index is available online.[4] Similarly, there were marginal changes proposed for the 2018

[3] We use a relatively weak correlation threshold of 0.3 given the relatively constrained degrees of freedom available for an Index measure in any given year (maximum n=50). See for example: Staiger D, Dimick JB, Baser O, Fan Z and Birkmeyer JD. Empirically derived composite measures of surgical performance. Medical Care 2009;47: 226- 233. Hays RD, Hayashi T. Beyond internal consistency reliability: rationale and user's guide for multi-trait analysis program on the microcomputer. Behavioral Research Methods 1990;22(2):167-75.

[4] *Methodology for the 2017 National Health Security Preparedness Index* <https://nhspi.org/wp-content/uploads/2017/04/NHSPI-2017-Methodology-PDF.pdf>.

Index release. Following the open call for measures and the public comment period on suggested changes that took place from August to November 2017, the total measures increased from 139 to 140 for all five years of data.

The Call for Measures for the 2019 release took place during July and August 2018. The Index program office received 22 comments about existing measures, four recommendations for new measures, and three general comments. These responses to the call for measures came from 13 different individuals who represent state health agencies, local emergency management operations, national associations, and federal health and regulatory agencies. The sections that follow reflect these changes as approved and adopted for the May 2019 release of the Index.

III. INDEX CONCEPTUAL FRAMEWORK AND STRUCTURE

The 2019 Index release retains the overall conceptual framework and structure developed for the original 2013 Index release by a broad collection of preparedness stakeholders. However, some changes to the subdomain structure have been implemented, as described below. The Index conceptual framework includes a set of six domains of health security activity that research and experience have shown to be important in protecting people from the health consequences of disasters and other large-scale hazards and emergencies:

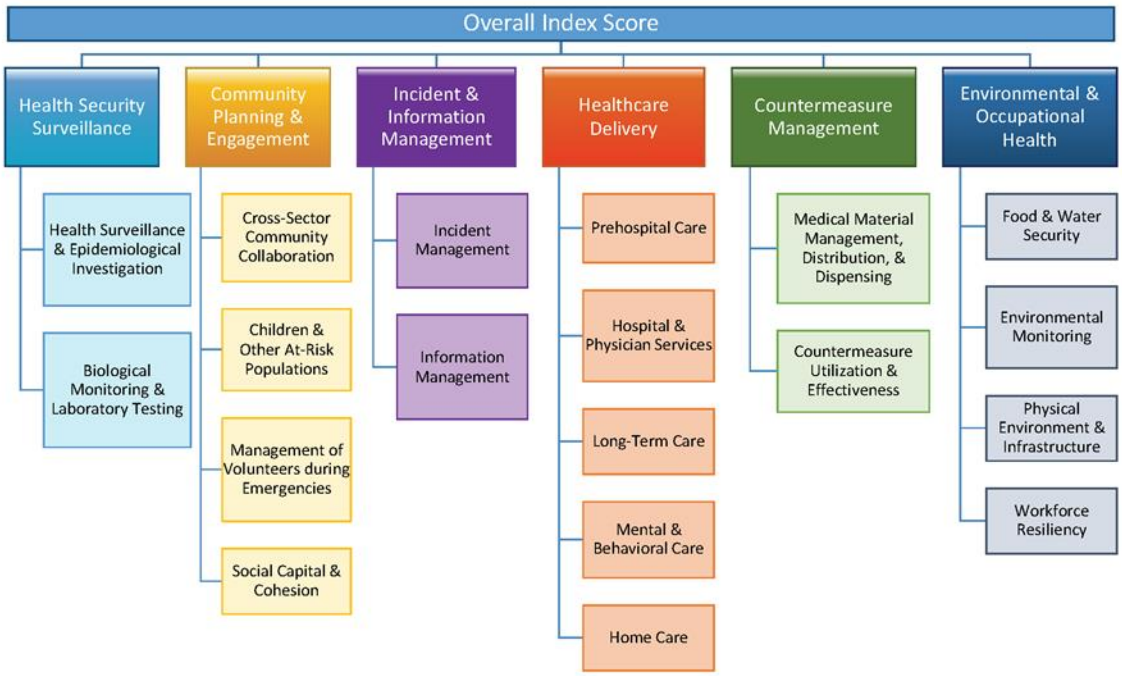
1. **Health security surveillance:** actions to monitor and detect health threats, and to identify where hazards start and spread so that they can be contained rapidly;
2. **Community planning and engagement:** actions to develop and maintain supportive relationships among government agencies, community organizations, and individual households; and to develop shared plans for responding to disasters and emergencies;
3. **Information and incident management:** actions to deploy people, supplies, money, and information to the locations where they are most effective in protecting health and safety;
4. **Healthcare delivery:** actions to ensure access to high-quality medical services across the continuum of care during and after disasters and emergencies;
5. **Countermeasure management:** actions to store and deploy medical and pharmaceutical products that prevent and treat the effects of hazardous substances and infectious diseases, including vaccines, prescription drugs, masks, gloves, and medical equipment; and
6. **Environmental and occupational health:** actions to maintain the security and safety of water and food supplies, to test for hazards and contaminants in the environment, and to protect workers and emergency responders from health hazards while on the job.

Each of these domains is further divided into a set of subdomains that reflect more specific areas of practice and policy. Finally, each subdomain contains individual measures that reflect specific preparedness capabilities. A total of 19 subdomains are operational in the 2019 Index release and each contains at least two, but in most cases multiple, related measures. Each measure is defined and constructed for a state and year, and standardized to a common scale using a methodology described below. Drawing from over 60 data sources, the 2019 Index release includes measure data for the 50 states as well as DC for the years 2013 through 2018.

There are no changes in the 2019 Index structure. However, the 2018 Index release subdomain structure was changed from the 2017 release in three domains: Incident and Information Management, Countermeasure Management, and Environmental and Occupational Health. This resulted in two subdomains in the Incident and

Information Management Domain instead of three: Incident Management and Information Management. With the removal of the Non-Pharmaceutical Intervention Subdomain, the Countermeasure Management Domain was left with two subdomains instead of three. And, the Environmental and Occupational Health Domain has four subdomains with the addition of two new subdomains: Physical Environment and Infrastructure Workforce Resiliency. With the consolidation, removal, and addition of these subdomains, the 2018 Index had a total of 19 subdomains within the six domains (see Figure 1)—which is unchanged with the 2019 Index release.

Figure 1: 2019 Index Release Domains and Subdomains



Index values are calculated using a series of sequential calculations that follow the Index structure:

- State-specific summary measures are constructed for each subdomain by computing a weighted average of the measures within each subdomain.
- State-specific summary measures are constructed for each domain by computing a weighted average of the subdomain summary measures.
- A state-specific overall Index summary measure is constructed by computing a weighted average of the domain summary measures.
- National summary measures are computed at the subdomain, domain, and overall Index level by computing a linear average of the 50 states and DC summary measures.

The use of a linear average of state measures rather than a population-weighted average in constructing national summary measures reflects a key principle of the Index conceptual framework: national health security requires that the nation be prepared to respond to disasters regardless of where they occur and how many people are affected. Consequently, the preparedness capabilities of each state and DC are defined to contribute equally to national health security.

IV. MEASURE SELECTION

The 2019 release of the Index is composed of scores for 129 individual measures, which include 19 Foundational Capability measures. The 2019 Index release builds on the methodology that was implemented for the 2016, 2017, and 2018 Index releases. Construction of the 2016 Index began with a pool of 197 individual measures identified by stakeholders involved in prior releases of the Index and used in the 2014 Index release.[5] For all measures used in the 2014 Index release, a series of measurement validity and reliability tests were used to remove redundant measures and measures lacking a strong empirical association with the Index domain and subdomain areas. Multi-trait analyses to examine the construct validity of each measure were conducted at both the domain and subdomain levels. These tests rely on the assumption that subgroups of measures related to a common underlying preparedness construct should demonstrate at least a moderate level of correspondence among measures. Based on the Central Limit Theorem of mathematics and statistics, we assume that combining unrelated measures in to summary measures at the subdomain and domain levels produces undesirable measurement noise that weakens the ability of Index summary measures to detect meaningful differences in health security levels across domains, across states, and over time. This assumption was tested empirically using multi-trait analysis.[6]

For these multi-trait analysis tests, a relatively weak correlation threshold of 0.3 was used given the relatively constrained degrees of freedom available for an Index measure in any given year (maximum n=50)[7], and given an assumption that individual preparedness capabilities within a domain or subdomain may exhibit relatively high heterogeneity and differentiation. Measures that fell below this threshold were considered to have low construct validity and contribute more noise than signal to Index summary measures.

Measures were recommended for exclusion from the 2016 Index if they showed evidence of low construct validity, including: (1) the measure performs poorly on construct validity tests at both the domain and subdomain level, as indicated by an adjusted multi-trait item-to-scale correlation coefficient of less than 0.3; (2) the measure's construct validity does not improve when reclassified into another domain or subdomain scale; and (3) the measure's validity and reliability has not been established through previously published studies. However, selected measures were retained in the Index despite low construct validity based on feedback obtained during the public comment period on proposed Index changes, and based on expert opinions expressed by members of the Index National Advisory Committee. A total of 28 measures were eliminated from the 2016 Index based on low construct validity, public comments, and expert judgment.

A final group of measures were classified as Foundational Capability measures because they reflect activities that are firmly ingrained in practice in all U.S. states and therefore do not vary across states or over time. As constants, these measures could not be evaluated empirically for their construct validity, and therefore were evaluated for Index inclusion solely based on expert opinions of members of the Index National Advisory Committee. The

[5] Please see *Methodology for the 2017 National Health Security Preparedness Index* <https://nhspi.org/wp-content/uploads/2017/04/NHSPI-2017-Methodology-PDF.pdf> for detailed information on the key differences between early versions of the Index and more recent versions.

[6] See Staiger D, Dimick JB, Baser O, Fan Z and Birkmeyer JD. Empirically derived composite measures of surgical performance. Medical Care 2009;47: 226- 233. Hays RD, Hayashi T. Beyond internal consistency reliability: ratio rationale and user's guide for multitrait analysis program on the microcomputer. Behavioral Research Methods 1990;22(2):167-75.

[7] The n=50 because this was prior to the District of Columbia being added to the Index.

measure weighting methodology, described below, was used to determine how to incorporate the Foundational Capability measures into Index summary measures.

The resulting 2016 Index release consisted of 134 individual measures, including a group of 18 Foundational Capability measures. It was from this framework of measures that we began the construction of the 2017 Index. Ultimately, we arrived at 139 individual measures for the 2017 Index release—which included 19 Foundational Capability measures. Following a similar approach outlined above, the 2018 Index release was based on the previous versions and included 140 measures, 19 of which are Foundational Capability measures. Similarly, resting on the previous work, the 2019 Index has 129 measures—which includes the 19 Foundational Capability measures.

V. MEASURE SCALING AND NORMALIZATION

The 2019 Index release uses a normalization method to convert each measure to a standardized scale before combining measures into subdomain, domain, and overall composite measures of preparedness. Normalization improves the validity and reliability of composite measures by placing component measures on a uniform scale before combining them.

The 2013 and 2014 releases of the Index used a normalization methodology for continuous measures that expressed each value as a proportion of the maximum value observed for that measure, after trimming (Winsorizing) any maximum values that exceed 2.5 standard deviations of the measure. In many cases, this method of normalization distorted significantly the distribution of the original measure because it did not incorporate information on the measure's variance or range into the scaling.

For the Index, this prior method of scaling had the additional, unintended effect of making dichotomous measures much more influential in the Index compared to continuous measures, because continuous measures are normalized to restricted ranges that are much less likely to contain values at or near zero.

To prevent these distortions in scaling, the 2019 Index release continues to use an alternative method of scaling implemented in 2016 that normalizes each measure to a common 0-1 range based on the full range of original data for all time periods from 2013 to 2017. This method, known as Min-Max scaling, calculates normalized values using a method that preserves the relationships among the original data values, as follows:

$$Standardized\ Value = \frac{(Original\ Value - Minimum\ Value)}{(Maximum\ Value - Minimum\ Value)}$$

where the minimum and maximum for each indicator are calculated across States and years to allow for longitudinal comparisons of Index values over time.

This method assigns each continuous and dichotomous variable to a common 0-1 scale based on the full range of observed values, placing both types of variables on equal footing when aggregating them into subdomain and domain measures.[8] This normalization method is like the z-score in its distributional properties but produces more

[8] Organization for Economic Cooperation and Development. Handbook on Constructing Composite Indicators. Paris: OECD and European Commission; 2008.

stable values than the z-score when used in small samples, as is the case with the Index's sample size of 50 states and DC.

After normalizing each measure using the min-max method, each measure is multiplied by 10 to place it on a 10-point scale with 10 being the highest possible level of preparedness.

The normalization method employed in the 2019 Index uses the maximum state value to define the upper boundary or "frontier" of preparedness for each measure. As such, this method defines "optimal" preparedness based on observed state values rather than based on a theoretical maximum value, an a priori goal, or an evidence-based standard derived from research. In the absence of a strong theoretical or evidence-based approach for defining the preparedness frontier, this "experience-based" method of scaling measures relative to the maximum state value provides a feasible approach to scaling that is commonly used in quality improvement applications.

VI. WEIGHTING

The 2013 and 2014 releases of the Index use the linear (unweighted) average as the method of aggregating individual measures into subdomain measures, aggregating subdomain measures into domain measures, and aggregating domain measures into the overall composite index of preparedness. This method implicitly assigns greater weight to measures located within subdomains and domains having fewer measures. As a result, large differences exist in the relative influence of each measure on overall Index results, and the most influential measures may not be the measures that are considered to be the most important to preparedness and national health security.

To address this distortion in implicit weighting, we used an expert panel methodology to develop and assign explicit weights to Index measures for use in constructing subdomain, domain, and overall preparedness composite measures. An online multi-stage Delphi process was used for this purpose, eliciting expert judgements about the importance of each measure, subdomain, and domain to national preparedness and health security.
In constructing the 2018 Index,[9] a separate expert panel was convened for each domain included in the Index model, with each panel comprised of around 20 subject matter experts who were identified through a nomination process and reviews of the preparedness scientific and professional literature. In total, 285 experts were identified and invited to participate in the item measure assessment, with 148 experts participating (52 percent). Two iterations of Delphi surveys and feedback reports were used with each panel to achieve convergence on expert ratings of importance. The convergence is facilitated by sharing the results from the initial assessment with each panel, and asking them to reassess the importance of a capability considering the collective intelligence of the entire panel. The iterative Delphi assessment of individual measures was conducted from December 5, 2017 to January 19, 2018.

The assessment of the 14 individual measures, grouped by the associated subdomains, included in the 2018 Index Delphi exercise used electronic surveys to elicit expert ratings of the importance of each measure to the capability construct reflected in each subdomain. A visual analog scale (VAS) was used to elicit expert ratings of importance on an interval scale (where "0" indicates that the measure is "not important at all" to the subdomain and "10"

---

[9] Please see *Methodology for the 2017 National Health Security Preparedness Index* <https://nhspi.org/wp-content/uploads/2017/04/NHSPI-2017-Methodology-PDF.pdf> for detailed information on previous Delphi methods and approaches.

indicates that it is "extremely important"), following methods that are well established for expert panel weighting processes.[10] Experts were also given the option of checking a box to indicate they "do not know" how important the measure is to the subdomain.

Similarly to assess the four subdomains, grouped by the associated domains, included in the 2018 Index Delphi exercise, expert panelists assessed the importance of each subdomain to the overarching capability construct reflected in the corresponding domain description. Instead of using a visual analogue scale for the subdomain assessments, experts were asked to allocate 100 percentage points for each domain across its subdomains. For each domain, panelists were asked to think about the core constructs within the domain description and then determine what proportion of a state's overall preparedness level is attributable to each of the subdomains making up that domain.

Likewise, the 2019 Index Delphi exercise was conducted in a similar fashion, with one notable difference—we did not divide the participants into separate panels. Each of the 283 experts was invited to assess each of the 10 items, and 129 experts participated (46 percent). The iterative Delphi assessment of individual measures was conducted from Feb. 11-March 6, 2019.

Delphi participants were given these instructions:

> The National Health Security Preparedness Index aims to provide an accurate portrayal of our country's health security using relevant, actionable information to help guide efforts to achieve a higher level of health security preparedness. While not an exhaustive compilation of national health security measures, the Index uses existing state-level preparedness data from a variety of sources that are broadly indicative of national or state-level health security. The current Index includes 140 measures grouped into 19 "sub-domains," which are organized into six "domains." In addition, we have identified 10 candidate measures and two new subdomains for possible inclusion in the next release of the Index.
>
> The goal of this Delphi exercise is to collect input on the relative importance of various factors affecting health security. Requiring between five and 10 minutes to complete, we have organized the Delphi so that a rater will assess a subset of the measures—not all of them. Using a visual analog scale that ranges from 0 to 10, raters will be presented with a list of one to three item measures and asked to assess how important each measure is to its overarching dimension of health security (or subdomain). Some raters will be asked to assess the importance of the proposed new subdomains. In that case, one will be asked to distribute 100 points between the subdomains to reflect their relative importance to the domain.
>
> When making assessments, one should think about the importance of an item measure across the full spectrum of the fifteen all-hazards National Planning Scenarios. These scenarios represent a broad range of potential health threats, and as noted in the National Health Security Strategy and Implementation Plan, "they can be intentional or naturally occurring and can result from both persistent and emerging threats, including severe weather, infectious diseases, hazardous material exposures, and terrorist attacks."

---

[10] Graham B1, Regehr G, Wright JG. Delphi as a method to establish consensus for diagnostic criteria. Journal of Clinical Epidemiology. 2003 Dec;56(12):1150-6.

> We recognize that some hazards are more relevant in some parts of the country than in others, and that some factors associated with health security might be more relevant to one planning scenario than another. For example, factors affecting health security preparedness in the context of Pandemic Influenza might be quite different from those in a Cyber Attack. An assessment of relative importance might be affected by the perceived likelihood of a scenario happening, its potential impact on health security should it happen, and the perceived relevance of a sub-domain for any given scenario. We ask that you think about these issues “on average.” That is, any given measure, sub-domain, or domain might have more relevance in the context of one planning scenario compared to another, but across all fifteen planning scenarios you should think about its “average” importance. Similarly, we ask that you make your assessments assuming an all-hazards approach—which is an approach for prevention, protection, mitigation, response, and recovery that addresses a full range of threats and hazards, including domestic terrorist attacks, natural and [human-caused] disasters, accidental disruptions, and other emergencies.

We assessed the degree of convergence among expert panelists in their assessments of the importance of each measure and subdomain using the discrimination coefficient of variation, which is superior to conventional measures of agreement such as the kappa statistic when used with measures on a continuous scale.[11] Using the conventional threshold of convergence of less than or equal to 0.5, all measures and subdomains were within the range of convergence (average = 0.29). After the second wave of the Delphi assessment, we concluded that sufficient agreement was attained among Delphi expert panelists concerning the importance of each measure to preparedness and health security constructs.

Weights for individual measures, subdomains, and domains were derived from the Delphi ratings by using the median VAS value for each individual measure and the median percentage value for each subdomain and domain. Weights were normalized within each subdomain such that they sum up to 1, by making the following transformation on the original weights:

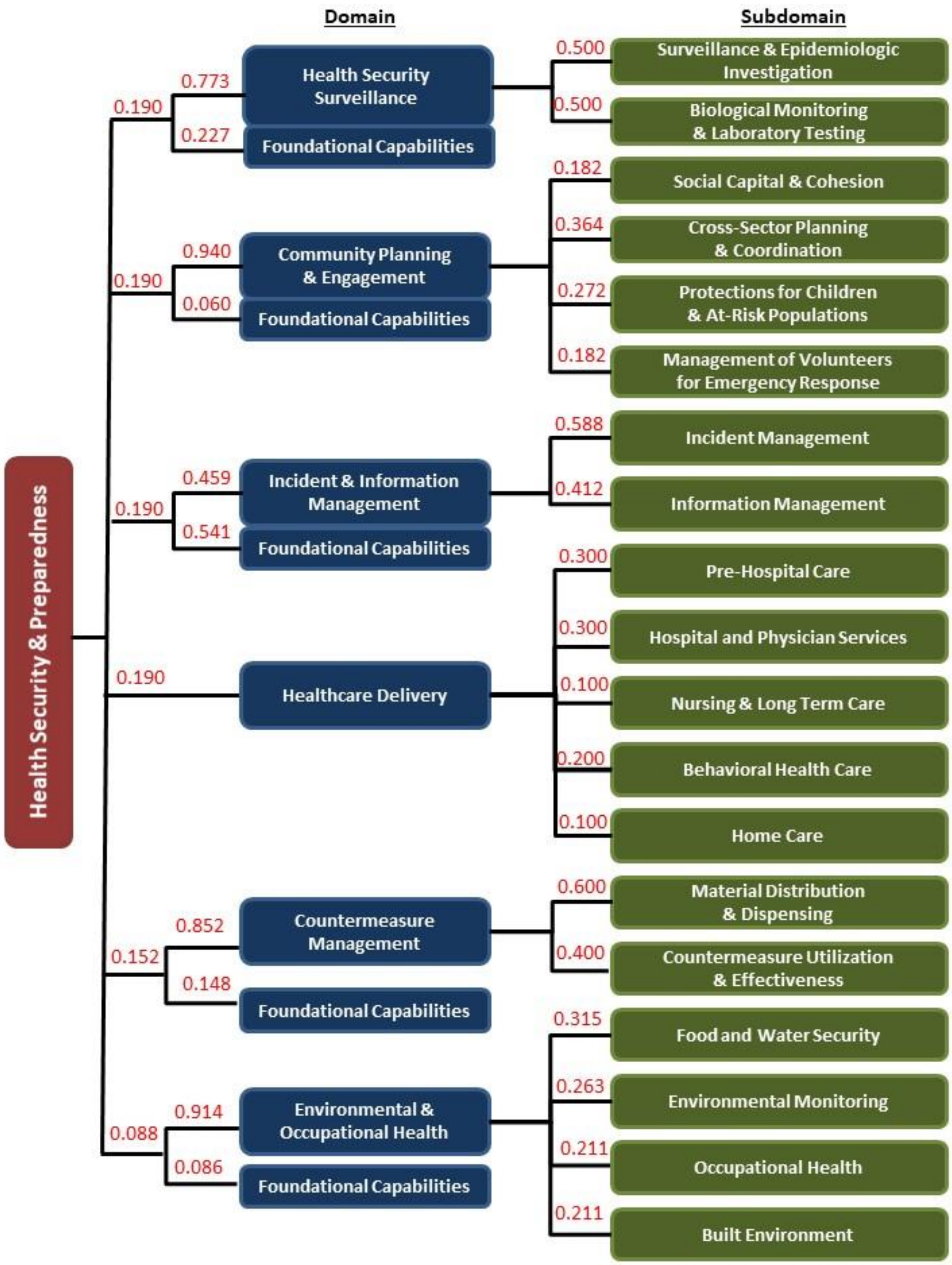
$$w_i' = \frac{w_i}{\sum_{j=1}^{n} w_j}$$

where n is the number of measures within each subdomain.

These weights were then used to calculate weighted averages as summary measures at the subdomain, domain, and overall Index levels for each state and each year. Foundational Capability measures were constructed as constants and averaged into the domain and overall summary measures using expert panel weights assigned to each measure defined as a Foundational Capability. Figure 2 shows the weights applied at the subdomain and domain levels in the 2019 Index release.

[11] Singh AC, Westlake M, Feder M, A generalization of the coefficient of variation with application to suppression of imprecise estimates. Journal of the American Statistical Association Section on Survey Research Methods 2011; 4359-4365.

Figure 2: Delphi Weights for Constructing Index Subdomain, Domain, and Overall Summary Measures



NOTE: numbers indicate Delphi expert panel weights

VII. IMPUTING MISSING VALUES

Some of the measures used in the Index have missing values for selected states due to incomplete response rates in the underlying data sources. The 2013 and 2014 releases of the Index address this missing values problem using a method that sets a missing value for a given measure and a given state equal to the unweighted average of that state's remaining measures in the same subdomain. This method introduces substantial measurement error into the Index, particularly given that most Index measures are not highly correlated at the subdomain level. This method also distorts the weighting system used within the Index by giving certain measures disproportionate influence on subdomain, domain, and overall Index values. A more accurate way of dealing with missing values is to use a statistical imputation method that predicts missing values using available information from a broader range of measures, including Index measures from all states and all available years.

In the 2019 Index measure set, 2.2 percent of the total item measures elements used for computing Index values are missing and must be imputed to generate Index summary measures. To reduce Index distortions due to missing values, we use multivariate regression models for each measure to predict its missing values, using five years of data on the measure along with other covariates in the model as predictors. Other covariates used in each regression model include non-missing Index measures from the same subdomain and domain, as well as state-level demographic, socioeconomic, and geographic characteristics, including population size, educational attainment, per capita income, age distribution, land area, population distribution across urban and rural areas, and climate region. We also ensure that no imputed values are greater than the maximum value of known data or conversely, less than the lowest value. Sensitivity analyses confirm that using imputed missing values derived from the regression models do not introduce any measurable bias in the Index summary measures constructed at the subdomain, domain, and overall Index levels.

VIII. AGGREGATING INDIVIDUAL MEASURES INTO SUMMARY MEASURES

The 2019 Index aggregates individual measures into subdomain, domain, and overall Index summary measures for each state and year using a weighted arithmetic mean. Since the Delphi weights for each measure are normalized as described above, the weighted mean is calculated as:

$$\bar{x} = \sum_{i=1}^{n} w_i' \, x_i$$

Where, x_i is the score for measure, subdomain, or domain i, and n is the number of measures, subdomains, or domains.

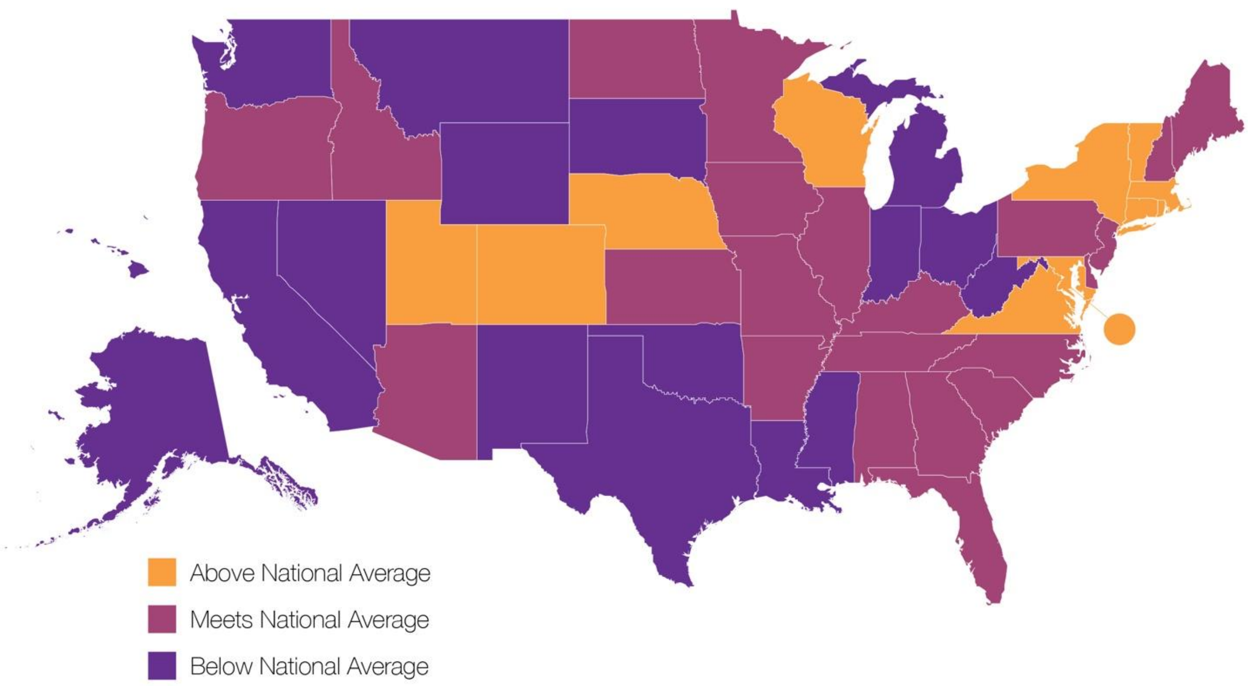
Summary measures are calculated for each state and year using this method. As final step, national summary measures are calculated at the subdomain, domain, and overall Index level for each year by calculating a linear (unweighted) mean of the 50 states and DC summary measures.

IX. CONFIDENCE INTERVALS FOR NATIONAL SUMMARY MEASURES

The 2013 and 2014 releases of the Index cautioned that the accuracy of comparisons made across domains, subdomains, and individual states had not been established. These previous Index releases suggested using a rule of thumb that differences of less than 10% may not be meaningful, but this rule of thumb is not based on any empirical results estimated from the Index data. Users wishing to use the Index for decision support and quality improvement require more robust and reliable information about the uncertainty surrounding Index measures and comparisons.

To address this need, the 2018 Index release continues the process initiated in 2016 and calculates confidence intervals for each of the national summary measures included in the Index at subdomain, domain, and overall levels. The confidence intervals allow summary measures for each state to be compared to the corresponding national summary measures. We use a conservative statistical threshold for Type I error of $p=0.01$ in calculating national confidence intervals, in order to account for the fact that state summary measures are also measured with error. As a result, the national confidence intervals used in the 2018 Index release represent 99% confidence intervals surrounding each national mean value. As shown in Figure 3 below, we are able to display the mean Index values for each state relative to these national confidence intervals, noting which states are significantly above or below the national average.

Figure 3: State Index Summary Measures Compared with National Averages



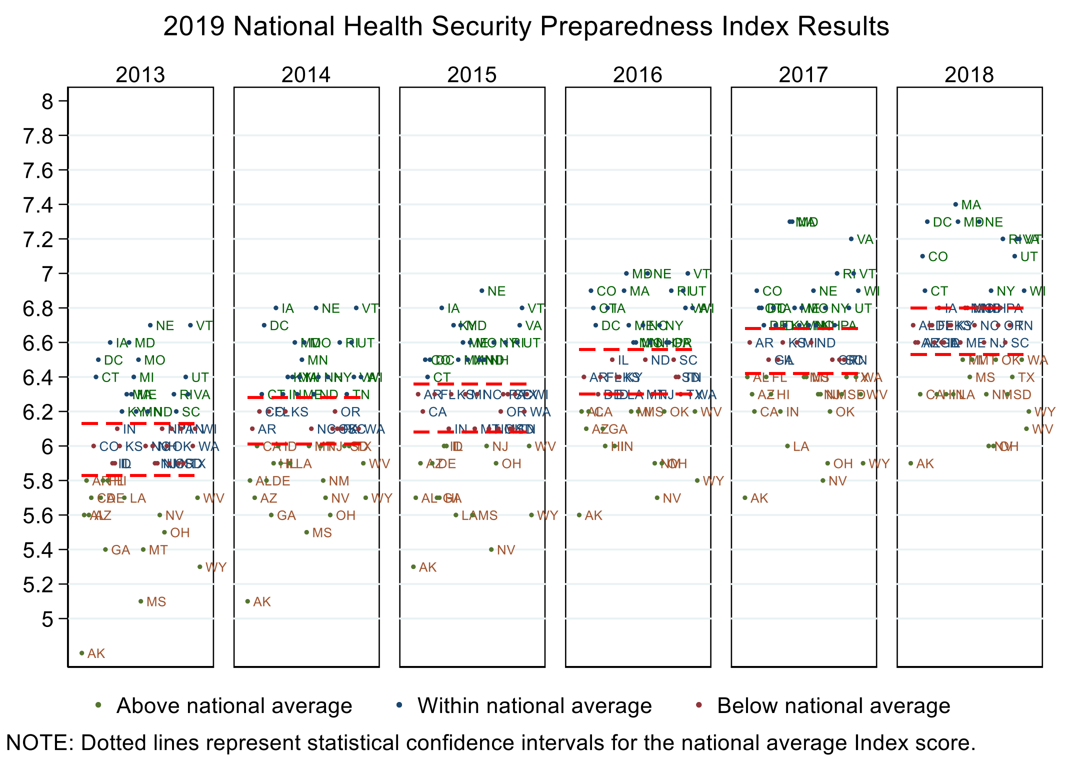
X. LONGITUDINAL COMPARISONS

The 2013 and 2014 releases of the Index did not support longitudinal comparisons of Index values over time (trending) due to significant differences in the measures and methodologies used in 2013 and 2014. The 2014 release cautioned users not to compare 2014 results with results found in the previous year's release. Modifications to the Index measures and methodologies are expected to occur with each annual release of the Index due to advances in preparedness science and due to changes in underlying data source availability, content, and quality. Nevertheless, if the Index is to become a valuable decision support tool in policy and practice communities, then users need to make valid comparisons of Index values over time and assess the direction and magnitude of change.

The 2019 Index release continues the process implemented in the 2016 release, and computes Index values not only for 2017, but also for 2013, 2014, 2015, and 2016, allowing for retrospective longitudinal comparisons each year dating back to the initial release year of 2013. Confidence intervals for the national average for mean Index values are also calculated at the subdomain, domain, and overall Index levels for all five annual periods, which makes it possible to track annual changes in the classification of each state's summary measures when compared to the national average.

It is important to note, however, that the time frame for each measure reflects the most recent data available for each year, which varies depending on the measure and its data source. One-year differences in Index values may be conservative estimates of change because the data for some measures are updated every two or three years rather than annually. Measures used in calculating Index values for 2017, for example, reflect the most recent time periods of data collection that range from calendar year 2013 to calendar year 2018. In constructing 2018 Index release values for 2016, 2015, 2014 and 2013, the time periods for individual measures are moved back by one, two, three, and four years, respectively, from the most recent year of data used in anchoring the 2017 data calculation. For measures that are collected every two or three years rather than annually, values are held constant during the time periods between data collection, in keeping with the policy of using the most recently available data for each time period. Consequently, estimates of longitudinal change in Index values should be considered conservative estimates of change.

Figure 4: Longitudinal Comparisons of State and National Index Values



XI. ACKNOWLEDGEMENTS

The National Health Security Preparedness Index is a program of the Robert Wood Johnson Foundation. The Program Office for the Index is based at the University of Kentucky and staffed through a collaboration between the Center for Public Health Systems and Services Research, College of Public Health and the Center for Business and Economic Research, Gatton College of Business and Economics.

Report Authors:

Glen P. Mays, PhD, MPH; Michael T. Childress, MA; Pierre Martin Dominique Zephyr, MS; Anna Goodman Hoover, PhD, MA; Nurlan Kussainov, MPP; Jeff Spradling, MA.

Recommended Citation:

Center for Public Health Systems and Services Research. Methodology for the 2019 Release, National Health Security Preparedness Index. Lexington, KY: University of Kentucky; May 2019.

Contributors at the Robert Wood Johnson Foundation:

Lori K. Grubstein, MPH, MSW, MPA, Program Officer

Alonzo Plough, PhD, MA, Vice President, Research-Evaluation-Learning and Chief Science Officer

National Advisory Committee Members, 2018-2019:

Thomas V. Inglesby, MD (Chair), Johns Hopkins University Center for Health Security

Anita Chandra, DrPH, RAND Corporation

Eric Holdeman, Emergency Management Consultant

Ana-Marie Jones, Interpro

Dara Lieberman, MPP, Trust for America's Health

Robert Mauskapf, MPA, Colonel, USMC (ret.), Director of Emergency Preparedness, Virginia Dept. of Health

Suzet McKinney, DrPH, MPH, Illinois Medical District Commission

F. Christy Music, DoD Liaison, Program Director, Health and Medical Policy, Department of Defense

Stephen Redd, MD, CDC Office of Public Health Preparedness & Response

Kevin Yeskey, MD, Deputy Assistant Secretary for Preparedness and Response

Program Consultants:

Christopher R. Bollinger, PhD, University of Kentucky

Index Workgroups:

This work would not have been possible without the input and feedback provided by voluntary members of the Index Analytic Methodology and Model Design Workgroup.

For More Information:

National Health Security Preparedness Index Program Office
Center for Public Health Systems and Services Research
University of Kentucky
111 Washington Avenue, Suite 201, Lexington, KY 40536
Telephone: 859-257-2912
Email: HealthSecurity@uky.edu
Web: www.nhspi.org