Projections of COVID-19 Hospitalizations and Deaths

Projections of COVID-19 Hospitalizations and Deaths

AUTHOR   Patrick Wiese, ASA
Lead Modeling Researcher
Society of Actuaries

A Survey of COVID-19 Outbreak Forecasts for the United States

Since the outbreak began in January, numerous research groups throughout the world have developed models to forecast the number of deaths that could result from COVID-19.[1] In addition, some of these models estimate the demand that the outbreak could place on hospital resources. Governments have factored outbreak forecasts into their decision-making. For example, reacting to forecasts of a sharp peak in COVID-19 cases that could overwhelm hospitals, many governments implemented social distancing measures intended to "flatten the curve", the goal being to keep daily demand for hospital services at a level below available daily resources.

Because COVID-19 is a new pathogen, model builders must contend with many unknowns, and these unknowns introduce significant uncertainty into forecasts. The scientific community has had little time to collect and process data to understand the risk posed by COVID-19. It is not surprising, therefore, that attempts to project the outbreak forward in time have resulted in a wide range of forecasts. Not only do these forecasts vary across models, but a single model may produce a broad range of forecasts, reflecting uncertainty in key assumptions that modelers express by performing sensitivity analyses. Moreover, as new data has become available during the course of the outbreak, some researchers have updated their key assumptions, resulting in revised forecasts that may differ substantially from prior forecasts generated by the same model.

To develop a better understanding of the evolving range of forecasts for the United States, Dr. Thomas McAndrew and Dr. Nicholas Reich of the University of Massachusetts Amherst have, on a weekly basis beginning in mid-February, surveyed experts who are separately engaged in efforts to model the outbreak.[2] The weekly survey tracks experts' forecasts for the peak month of the outbreak, as well as their short-term and long-term estimates for the cumulative number of cases and deaths. Survey results are published in a weekly report entitled "COVID19-Expert Survey".[3]

The latest report from McAndrew and Reich, issued on April 8, reveals that about 60% of surveyed experts anticipate that peak hospitalization in the U.S. will occur in either April or May, while 40% estimate that the peak will occur sometime between June and August (Figure 1).[4] Experts were also asked to provide one-week and two-week forecasts for the cumulative number of COVID-19 cases in the U.S. Even across this relatively short forecasting period, the range of forecasts was quite broad (Figure 2).

---

[1] Some research groups have developed COVID-19 projection models from the ground up, while others have relied upon models developed for previous pandemics (such as the Swine Flu of 2009), adjusting model parameters to fit what is known about COVID-19.

[2] According to the report, "participants are modeling experts and researchers who have spent a substantial amount of time in their professional career designing, building, and/or interpreting models to explain and understand infectious disease dynamics and/or the associated policy implications in human populations."

[3] https://works.bepress.com/mcandrew/5/

[4] A total of 20 experts provided responses for the April 8th report. Survey questions were submitted to the experts on April 6, and the report was published 2 days later.

Figure 1
DISTRIBUTION ACROSS SURVEYED EXPERTS:
PROJECTED PEAK MONTH OF COVID-19 HOSPITALIZATIONS IN THE UNITED STATES
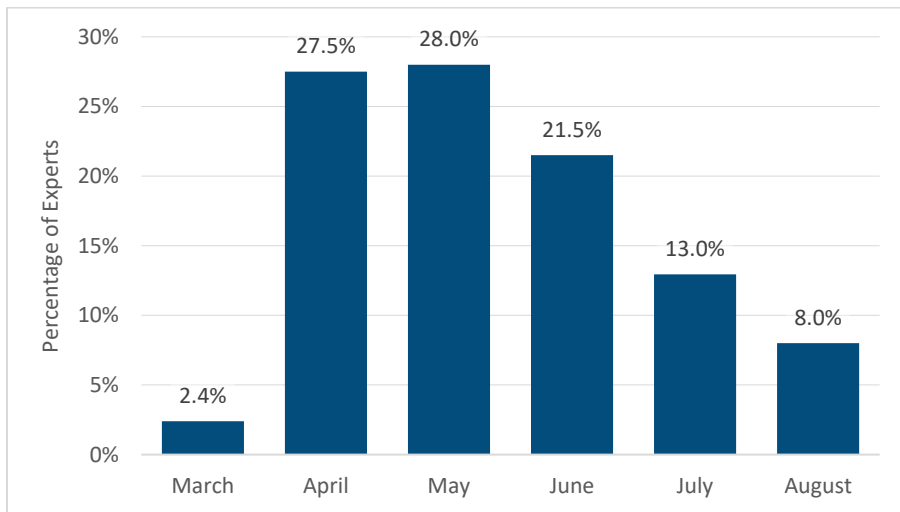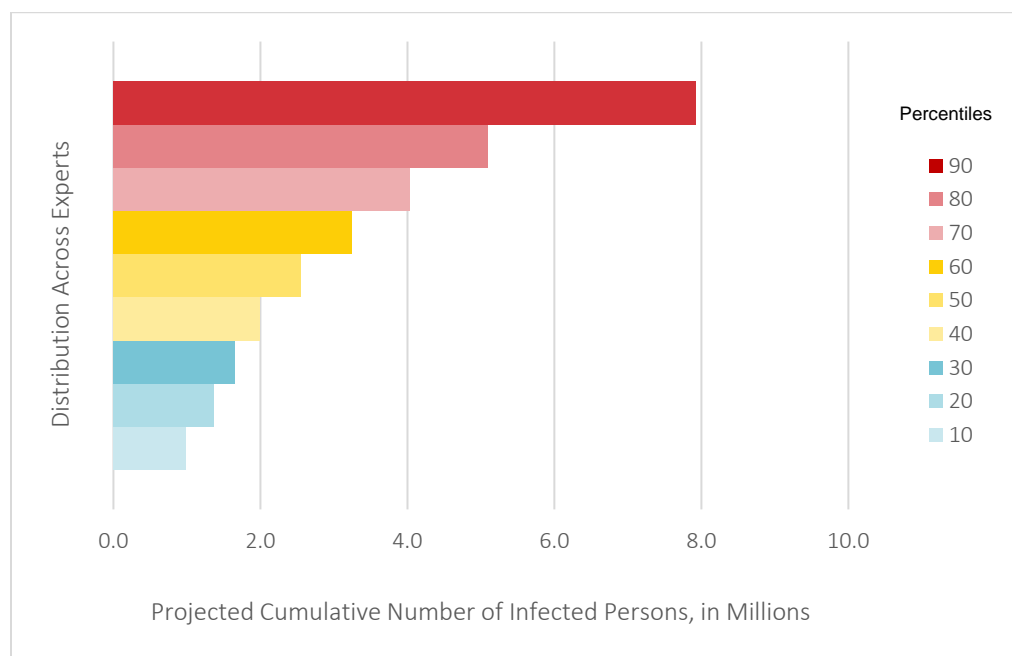

Source: COVID 19-Expert Survey-7-20200408

Figure 2
DISTRIBUTION ACROSS SURVEYED EXPERTS:
SHORT-RANGE PROJECTIONS OF THE CUMULATIVE NUMBER OF COVID-19 CASES IN THE UNITED STATES,
USING DATA FROM APRIL 6 AS THE STARTING POINT FOR FORECASTING
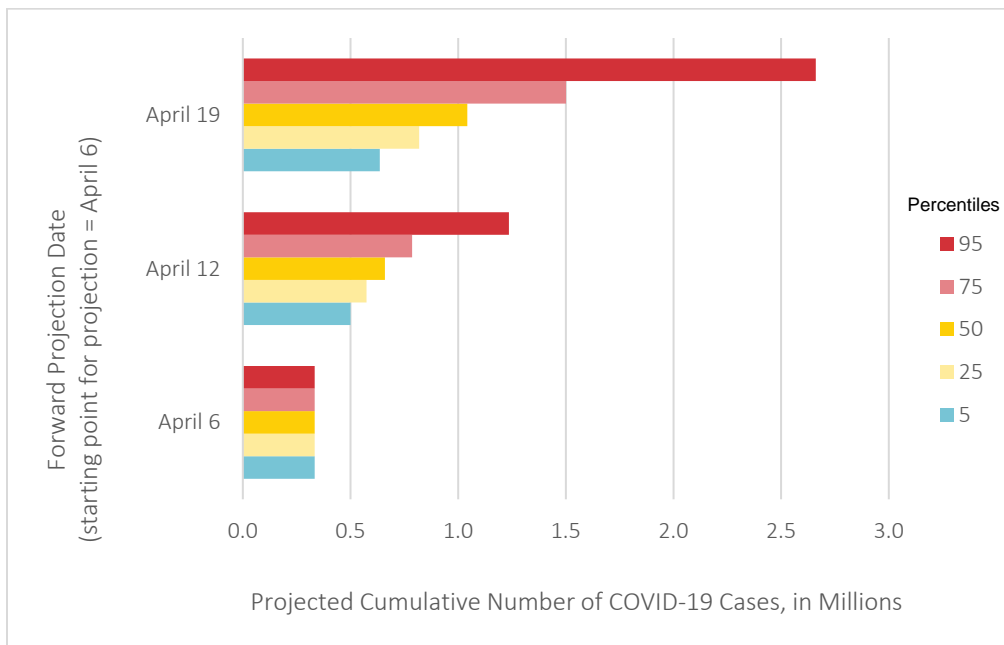
Source: COVID 19-Expert Survey-7-20200408

The survey not only reveals a broad distribution of outbreak forecasts, but also a wide range of opinions regarding the present status of the outbreak. For example, experts were asked to provide low, most likely and high estimates for the total number of Americans that have been infected with the virus through April 5 (one day prior to the date of the survey), including both symptomatic and asymptomatic persons. The experts' responses were merged into a consensus probability distribution. The 10th, 50th and 90th percentiles of this distribution are 1 million, 2.6 million and 7.9 million, respectively (Figure 3). This high level of uncertainty is due, in part, to the possibility that the virus might have little effect on some infected individuals, resulting either in mild symptoms or in no symptoms whatsoever. Without random testing of the general population, it is difficult to ascertain the size of this group of fortunate individuals. Yet it is a critical piece of information for understanding the risk characteristics of the virus: the greater the number of infected individuals who experienced little or no discomfort as a result of the virus – relative to the number who experienced severe complications or death – the less danger the virus poses to the population that has not yet been infected.

Figure 3

DISTRIBUTION ACROSS SURVEYED EXPERTS:
ESTIMATED CUMULATIVE NUMBER OF AMERICANS INFECTED BY CORONAVIRUS AS OF APRIL 5,
INCLUDING INDIVIDUALS WHO ARE ASYMPTOMATIC
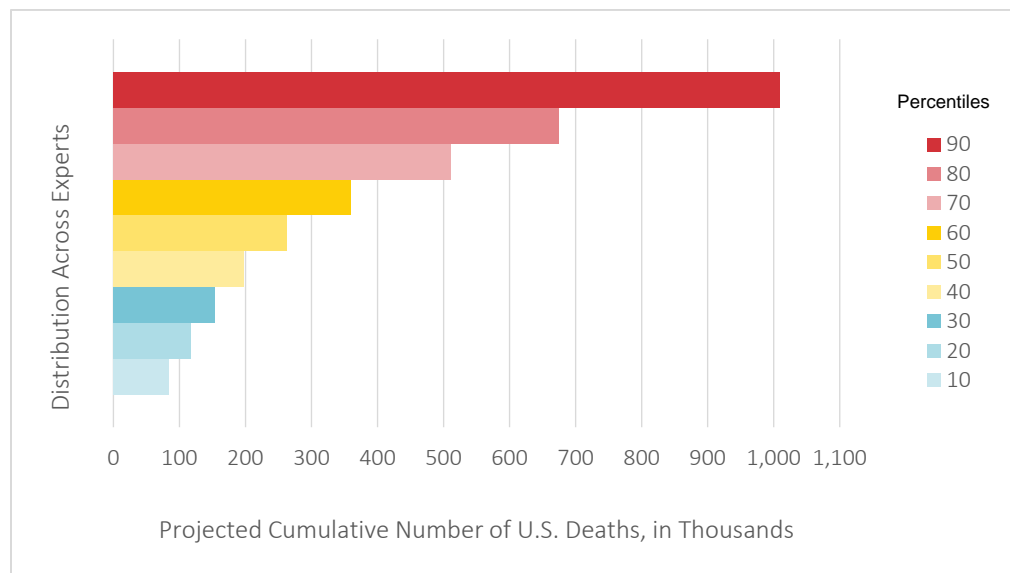

Source: COVID 19-Expert Survey-7-20200408

The McAndrew/Reich survey also tracks long-range outbreak forecasts extending through the end of 2020.[5] While no long-range questions were posed in the most recent weekly survey, the survey conducted the prior week included the following question: "what are the smallest, most likely, and largest number of deaths [in the United States] due to COVID-19 in 2020?"[6] The responses were merged into a consensus probability distribution (Figure 4).

---

[5] "Long-range" has a different meaning in the context of different types of models. When modeling weather, for example, the National Weather Service characterizes forecasts of up to one week in the future as short-range, while forecasts that look beyond one week into the future are considered long-range. This division of time arises because, roughly speaking, the ability to accurately predict weather declines significantly for forecasting horizons that exceed one week. A very different definition of short-range is used for projecting the Social Security Program (OASDI) forward in time. The annual OASDI Trustees Report defines short-range as up to ten years into the future, while long-range projections extend 75 years into the future.

[6] This survey was submitted to experts on March 30, and results were published on April 1.

The 10th, 50th and 90th percentiles of this distribution are 83.5 thousand, 262.5 thousand, and 1 million, respectively, and the average is 566 thousand.

Figure 4

DISTRIBUTION ACROSS SURVEYED EXPERTS:
ESTIMATES OF THE CUMULATIVE NUMBER OF COVID-19-RELATED DEATHS IN THE UNITED STATES
BY THE END OF 2020

Source: Consensus forecast from the COVID 19-Expert Survey-7-20200401

The Challenges of Modeling an Outbreak Caused by a Novel Pathogen

Recent history can provide some sense of the challenge of quantifying the risk posed by a new pathogen. In the spring of 2009, a novel influenza virus emerged and spread rapidly throughout the world. Referred to as "Swine Flu", researchers scrambled to process emerging data and develop estimates of the danger posed by this virus. In a report published in 2013, four years after the outbreak, the National Institute of Health (NIH) reviewed 77 estimates of Swine Flu case fatality risk from 50 published studies.[7] Many of the 77 estimates were produced in the first nine months of the pandemic. The NIH found "substantial heterogeneity in the published estimates, ranging from less than 1 to more than 10,000 deaths per 100,000 cases or infections." The report concludes that "our review highlights the difficulty in estimating the seriousness of infection with a novel influenza virus".[8]

With respect to SARS-CoV-2, the coronavirus that causes the illness known as COVID-19, questions abound regarding its transmission rate, the risk of death for those who are infected, the influence of comorbidities, the length of the infectious period, the number of people who have been infected but who have experienced no symptoms or mild symptoms that didn't require a trip to a doctor, how long the virus can survive on a surface, whether immunity is acquired by those who have been infected, and the influence of country, region and city-

---

[7] Wong, Jessica Y., et al. (2013, November 24). "Case Fatality Risk of Influenza A(H1N1pdm09): A Systematic Review." *Epidemiology (Cambridge, Mass.), 24(6),* 830–841. *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3809029/.*

[8] With the passage of time and the expansion of available data, researchers were able to produce increasingly reliable estimates of the mortality risk posed by (H1N1)pdm09, the virus that caused the Swine Flu outbreak of 2009. Today, more than ten years after the outbreak, the Centers for Disease Control and Prevention (CDC) estimates that 61 million Americans were infected with the virus during the one-year period beginning in April 2009, of which 12.5 thousand persons died. This translates into an infection mortality rate of 0.02%. This is less than the CDC's estimate of 0.1% for the mortality rate associated with seasonal flu.

specific factors on both the rate of transmission and the case fatality rate. Along with these variables, outbreak modelers must estimate the impact of social distancing measures intended to alter the trajectory of the outbreak, as well as assess the capacity of the health care system(s) to absorb the projected case load.

Even if a simulation model correctly predicts the short-range trajectory of an outbreak, this doesn't necessarily imply that the model's key assumptions are correct, and, as a consequence, the model's medium and long-range projections could prove to be off-the-mark. In a paper released on March 26 by the University of Oxford's Department of Zoology, using data from the United Kingdom and Italy, researchers demonstrated that the time series of COVID-19 deaths from late January through mid-March can be replicated (via simulation) using either an assumed infection fatality rate (IFR) of 1%, or a much lower assumption of 0.1%.[9] To replicate the UK and Italian time series of deaths using either an IFR of 0.1% or 1%, the researchers slightly varied the estimated starting date for the outbreak: the lower the assumed IFR, the earlier the estimated starting date for the outbreak, and the larger the estimated number of individuals who have, to date, been exposed to the virus and thus acquired immunity.

Obviously, it matters a great deal whether the IFR is 0.1% or 1% (or some other value). While either assumption is plausible in the view of the authors, and either assumption can be reconciled with January through March COVID-19 data, the long-range implications of these two assumptions are vastly different. An IFR of 0.1% is equal to that of the seasonal flu, and therefore represents a routine risk-level that society has successfully managed year after year. In contrast, an IFR of 1% is an order of magnitude greater, the equivalent of 10 seasonal flu seasons compressed into single year. Given that seasonal flu typically kills anywhere from 20 to 60 thousand Americans a year[10], and 300 to 650 thousand persons worldwide[11], a virus with an IFR of 1%, all else equal[12], could potentially kill hundreds of thousands of Americans and several million persons worldwide.

To reduce the level of uncertainty associated with IFR estimates for COVID-19, the Oxford paper argues that serological surveys are needed to determine the percentage of the general population that has already been exposed to the virus. The greater this percentage, the lower must be the IFR. However, most countries do not yet have the capacity to randomly test the general population; instead, testing has been focused almost exclusively on individuals who exhibit severe symptoms. Until testing can be expanded to cover the general population, the Oxford paper implies that IFR estimates will be fraught with uncertainty, and, consequently, the level of risk posed by COVID-19 will be difficult to ascertain.

Two Modeling Approaches for Projecting Outbreaks Forward in Time

Because policymakers, governments and citizens require some sense of the danger posed by a new pathogen, outbreak simulation models are necessary despite the fact that they can be difficult to construct and calibrate due to data limitations, and despite the fact that their forecasts may be subject to significant uncertainty. Of the COVID-19 models that researchers have shared with the public and described in online reports, there are two main types:

---

[9] Lourenco, Jose, et al. (2020, March 26). "Fundamental Principles of Epidemic Spread Highlight the Immediate Need for Large-Scale Serological Surveys to Assess the Stage of the SARS-CoV-2 Epidemic." MedRxiv. *https://www.medrxiv.org/content/10.1101/2020.03.24.20042291v1*.

[10] Centers for Disease Control and Prevention. "Past Seasons Estimated Influenza Disease Burden." *https://www.cdc.gov/flu/about/burden/past-seasons.html*.

[11] Centers for Disease Control and Prevention. "Seasonal Flu Death Estimates Increase Worldwide." *https://www.cdc.gov/media/releases/2017/p1213-flu-death-estimate.html*.

[12] The IFR is just one of the various key metrics that is used to quantify the threat posed by a virus. Another key metric is the "basic reproduction number" or "R0". The greater the value of R0, the greater the rate of virus transmission from person to person. In various studies, the R0 of seasonal flu has been estimated in the neighborhood of 1.3. In contrast, the R0 for COVID-19 has been estimated in the range of 2 to 3, indicating that it is more easily transmitted than the seasonal flu.

1) statistical models[13] and (2) mechanistic models. The Institute for Health Metrics and Evaluation (IHME) model[14], which has been frequently cited by the media as well as by the White House[15], is an example of a statistical model[16], while the Imperial College of London's model[17,18] is an example of a susceptible-infected-recovered (SIR) model, which is a type of mechanistic model.

A statistical model uses correlations or patterns in data to forecast the propagation of a virus. A common approach is to focus on the time series of virus-related deaths, separately by city or geographic region, fitting this data to a curve the describes the anticipated rise, peak and fall of the number of daily deaths. The curve is extracted from cities or regions that have already passed through the outbreak, such as Wuhan, China. The assumption is that, in each different region, the outbreak will follow a similar "shape", curve or pattern across time. A model may tweak or adjust the assumed outbreak shape to account for region or city-specific factors, such as delays associated with implementing social distancing measures.

In contrast to statistical models, mechanistic models focus on the dynamic processes through which a virus propagates through a population. Estimates of the transmissibility and lethality of the virus are used to simulate the progression of an outbreak across time. A SIR model, for example, projects shifts in the population from "susceptible" (i.e. not yet infected) to "infected", and from "infected" to either "recovered" or deceased. Some SIR models are quite simple, assuming that all persons have an equal chance of becoming sick, that infected persons are equally likely to transmit the virus, and that infected persons share the same probability of death. More complicated SIR models subdivide the population into groups, each group having distinct characteristics with respect to risk of infection, risk of transmission, and risk of death. Some SIR models go a step further, using an agent-based method to simulate unique individuals (as opposed to groups of individuals), each interacting with other unique simulated individuals.

To project hospital visits, models use data to assess the probability that an infected person will become seriously ill. Some models, such as the IHME's model, assume that hospital services are a function of the projected number of deaths. Many models not only forecast the demand for hospital services, but also compare that demand to available hospital beds in each geographic region, thereby developing a sense of where hospital strain is most likely to occur.

Relative to statistical models, SIR models are generally more "data hungry", but they provide greater modeling flexibility. For example, a SIR model's key parameters can be sensitivity-tested to produce a range of simulated outcomes, such as low, most-likely and high estimates for the number of hospitalizations and deaths. In addition, the impact of social distancing measures can be estimated by adjusting the basic reproduction number ("R0") which describes the rate of virus transmission.

Both statistical models and mechanistic models can quickly become "stale" during the early stages of an outbreak. With little data to draw upon, initial modeling efforts necessitate the use of assumptions that have a wide range of uncertainty. As an outbreak progresses, the pool of available data expands, providing researchers with valuable

---

[13] An earlier version of this report used the term "outbreak curve model". However, feedback received from several epidemiologists suggests that "statistical model" is the preferred term.

[14] Murray, Christopher JL. (2020, March 30). "Forecasting COVID-19 Impact on Hospital Bed-Days, ICU-Days, Ventilator-Days and Deaths by U.S. State in the Next Four Months." MedRxiv. *https://www.medrxiv.org/content/10.1101/2020.03.27.20043752v1.*

[15] Aizenman, Nurith. (2020, April 1). "Five Key Facts Not Explained in White House COVID-19 Projections." NPR. *https://www.npr.org/sections/health-shots/2020/04/01/824744490/5-key-facts-the-white-house-isnt-saying-about-their-covid-19-projections.*

[16] While the IMHE model uses a statistical approach to project the number of deaths, the component of the model that projects hospital service utilization is best described as mechanistic. Thus, the IMHE model has both statistical and mechanistic components.

[17] Flaxman, Seth, et al. (2020, March 30). "Report 13—Estimating the Number of Infections and the Impact of Non-Pharmaceutical Interventions on COVID-19 in 11 European Countries." Imperial College London. *https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-13-europe-npi-impact/.*

[18] Adam, David. (2020, April 2). "Special Report: The Simulations Driving the World's Response to COVID-19." Nature. *https://www.nature.com/articles/d41586-020-01003-6.*

information that can be used to revise their models.

Inevitably, model revisions result in shifts in outbreak forecasts. Large shifts could potentially undermine the public's faith in a model. However, revisions of forecasts do not, in general, arise from a lack of modeling expertise, but rather from data limitations that are part and parcel of dealing with a new pathogen. Revisions to forecasts are a sign that modelers are paying attention to the continuous influx of new data produced by researchers around the world, and diligently adjusting their models to reflect the most current available information about the pathogen.

In the coming weeks, this report will be periodically updated, providing a summary of the ongoing efforts of modelers to forecast the course of the outbreak.

About The Society of Actuaries

With roots dating back to 1889, the Society of Actuaries (SOA) is the world's largest actuarial professional organizations with more than 31,000 members. Through research and education, the SOA's mission is to advance actuarial knowledge and to enhance the ability of actuaries to provide expert advice and relevant solutions for financial, business and societal challenges. The SOA's vision is for actuaries to be the leading professionals in the measurement and management of risk.

The SOA supports actuaries and advances knowledge through research and education. As part of its work, the SOA seeks to inform public policy development and public understanding through research. The SOA aspires to be a trusted source of objective, data-driven research and analysis with an actuarial perspective for its members, industry, policymakers and the public. This distinct perspective comes from the SOA as an association of actuaries, who have a rigorous formal education and direct experience as practitioners as they perform applied research. The SOA also welcomes the opportunity to partner with other organizations in our work where appropriate.

The SOA has a history of working with public policymakers and regulators in developing historical experience studies and projection techniques as well as individual reports on health care, retirement and other topics. The SOA's research is intended to aid the work of policymakers and regulators and follow certain core principles:

Objectivity: The SOA's research informs and provides analysis that can be relied upon by other individuals or organizations involved in public policy discussions. The SOA does not take advocacy positions or lobby specific policy proposals.

Quality: The SOA aspires to the highest ethical and quality standards in all of its research and analysis. Our research process is overseen by experienced actuaries and nonactuaries from a range of industry sectors and organizations. A rigorous peer-review process ensures the quality and integrity of our work.

Relevance: The SOA provides timely research on public policy issues. Our research advances actuarial knowledge while providing critical insights on key policy issues, and thereby provides value to stakeholders and decision makers.

Quantification: The SOA leverages the diverse skill sets of actuaries to provide research and findings that are driven by the best available data and methods. Actuaries use detailed modeling to analyze financial risk and provide distinct insight and quantification. Further, actuarial standards require transparency and the disclosure of the assumptions and analytic approach underlying the work.

Society of Actuaries
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org