New Resource Adequacy Criteria for the Energy Transition

MODERNIZING RELIABILITY REQUIREMENTS

A Report by the
Energy Systems Integration Group's
Resource Adequacy Task Force

March 2024

ESIG
ENERGY SYSTEMS
INTEGRATION GROUP

ESIG
ENERGY SYSTEMS
INTEGRATION GROUP

About ESIG

The Energy Systems Integration Group is a nonprofit organization that marshals the expertise of the electricity industry's technical community to support grid transformation and energy systems integration and operation. More information is available at https://www.esig.energy.

ESIG's Publications Available Online

This report is available at https://www.esig.energy/new-resource-adequacy-criteria. All ESIG publications can be found at https://www.esig.energy/reports-briefs.

Get in Touch

To learn more about the topics discussed in this report or for more information about the Energy Systems Integration Group, please send an email to info@esig.energy.

New Resource Adequacy Criteria for the Energy Transition: Modernizing Reliability Requirements

A Report by the Energy Systems Integration Group's Resource Adequacy Task Force

Prepared By

Derek Stenclik, Telos Energy

Project Team

James Okullo, Energy Systems Integration Group

Aidan Tuohy, EPRI

Bethany Frew, National Renewable Energy Laboratory

Gord Stephen, National Renewable Energy Laboratory

Juan Pablo Caravallo, U.S. Department of Energy,
Office of Energy Efficiency and Renewable Energy

Task Force

Dan Arber, Australian Energy Market Operator

Justin Brooks, Duke Energy

Brent Buffington, Southern California Edison

Allison Campbell, Pacific Northwest National Laboratory

Casey Cathey, Southwest Power Pool

Aditya Choukulkar, Form Energy

Jon Cook, Salt River Project

Genevieve de Mijolla, EPRI

Chris Dent, University of Edinburgh

Andrew Dobbie, National Grid ESO

John Fazio, Independent Consultant

Frank Felder, Independent Consultant

Armando Figueroa-Acevedo, Midcontinent Independent System Operator

Jimmy Glotfelty, Public Utility Commission of Texas

Hwikwon Ham, Minnesota Public Utilities Commission

Will Harropp, Grid United

Charles Hendrix, Southwest Power Pool

Dor Hirsh Bar Gai, Northwest Power and Conservation Council

Daniel Hua, Northwest Power and Conservation Council

Yvonne Huang, New York Independent System Operator

Eduardo Ibanez, Midcontinent Independent System Operator

Julie Jin, Electric Reliability Council of Texas

Andreas Klaube, Northeast Power Coordinating Council

William Lamana, North American Electric Reliability Corporation

Eamonn Lannoye, EPRI

Christopher Lau, Hawaiian Electric Company

Stephanie Lenhart, Boise State University

Jacob Mays, Cornell University

Nils Muller, Amprion

Mark Olson, North American Electric Reliability Corporation

Keith Parks, Xcel Energy

Ralph Pfeiffer, Amprion

Rajat Pungaliya, Clearway Energy Group

Aaron Schwartz, Rocky Mountain Institute

This report was produced by a task force made up of diverse members with diverse viewpoints and levels of participation. Specific statements may not necessarily represent a consensus among all participants.

Suggested Citation

Energy Systems Integration Group. 2024. *New Resource Adequacy Criteria for the Energy Transition: Modernizing Reliability Requirements.* Reston, VA. https://www.esig.energy/new-resource-adequacy-criteria.

The work described in this study was funded by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy under the National Renewable Energy Laboratory Subcontract No. LAT-9-92162-01.

Contents

Abbreviations Used

CONE	Cost of new entry
CVaR	Conditional value at risk
EENS	Expected energy not served
ERCOT	Electric Reliability Council of Texas
EUE	Expected unserved energy
LOLE	Loss-of-load expectation
LOLEv	Loss-of-load events
LOLH	Loss-of-load hours
LOLP	Loss-of-load probability
MISO	Midcontinent Independent System Operator
NEUE	Normalized expected unserved energy
NWPCC	Northwest Power and Conservation Council
RA	Resource adequacy
SPP	Southwest Power Pool
VaR	Value at risk
VoLL	Value of lost load

Executive Summary

The transition toward a cleaner and more weather-dependent power system brings with it unprecedented challenges and opportunities for maintaining resource adequacy—namely, a future that includes rapid load growth, plant retirements, and a shift toward variable and energy-limited resources. Resource adequacy analyses, capacity accreditation, and the resource adequacy planning criteria are becoming increasingly important for power system planning and investment decisions.

While previous ESIG reports evaluated the changing resource adequacy risks and planning practices created by the energy transition, none specifically examined potential changes to the resource adequacy criterion used for planning. The resource adequacy criterion sets the level of supply- and demand-side resources that are required for a given power system to meet reliability objectives. It is a pivotal standard that influences billions of dollars of investment decisions.

Limitations of the Current Resource Adequacy Criterion

The most common resource adequacy criterion is the loss-of-load expectation (LOLE), and is colloquially referred to as the 1-day-in-10 LOLE criterion across much of North America. The LOLE criterion has been the lynchpin of power system planning for decades. But resource adequacy criteria now need to do more—being not only about capacity shortfalls at any one point in time, but also about energy constraints that arise from variable renewables, increasing battery storage, and limited fuel supplies. Damages from outages increase at a nonlinear rate as events increase in duration and size, so it is important to differentiate between large or long-lasting outages versus short ones, and consider tail risks associated with high-impact, low-probability events.

The LOLE criterion has several limitations. First, it is treated as an arbitrary line in the sand rather than articulating resource adequacy as a continuum and a trade-off between cost and reliability. Second, it does not differentiate types of shortfalls, but rather treats risks associated with longer-duration or larger outages as equal to shorter, less severe outages. Third, it is a static criterion which in many regions has not changed in decades, despite rapid changes to the power system resource mix and electrification of new sectors. Lastly, the minimum threshold for the criterion is often set without considering the trade-off between cost and reliability.

These limitations highlight a need to move beyond a single, one-size-fits-all resource adequacy criterion and augment it with multi-metric criteria. The ESIG Resource Adequacy Task Force identified four critical features for a new resource adequacy criteria:

- Measures the magnitude (maximum MW and total MWh) of energy shortfalls and not just the number of times that shortfalls occur (their frequency)
- Captures tail risks and outlier events
- Explicitly considers the inherent trade-off between cost and reliability

There is a need to move beyond a single, one-size-fits-all resource adequacy criterion and move toward multi-metric criteria.

Transitioning to Multi-metric Criteria

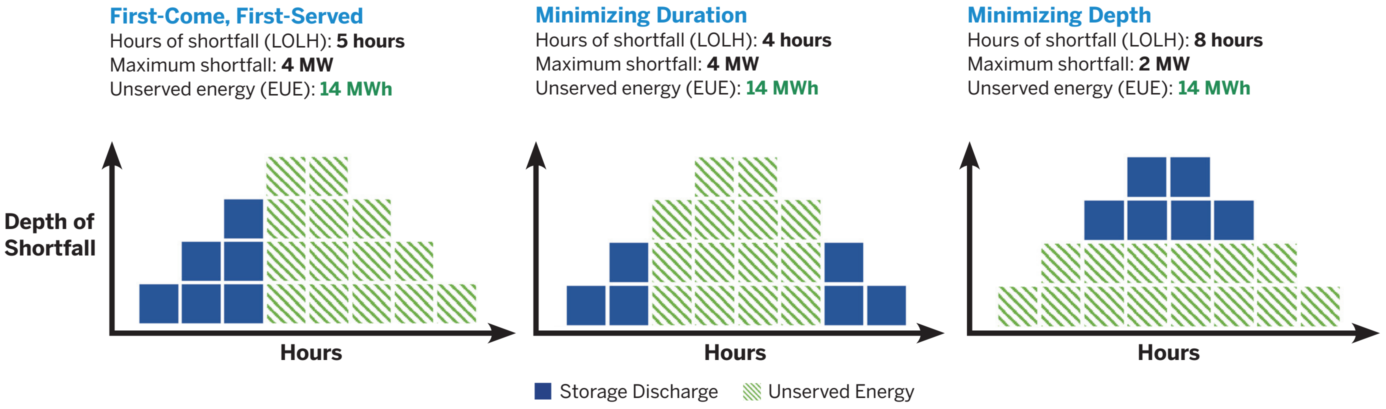
Loss-of-load expectation as the sole resource adequacy criterion represents only a single dimension of risk. It needs to be supplemented (p. 6). A significant limitation of the single criterion approach is its failure to differentiate among the size, frequency, duration, and timing of shortfalls. This is a critical omission, as damages associated with power system shortfalls are nonlinear. Longer and larger disruptions lead to disproportionately greater damages, yet the LOLE metric treats all resource adequacy shortfalls equally. This equal weighting does not accurately reflect the real-world impacts of loss of load, which vary greatly in severity and consequences. Tail events—those which may occur seldomly but have disproportionately high impacts and costs—require additional focus. Using resource adequacy criteria that explicitly differentiate between resource adequacy shortfalls is beneficial.

A significant limitation of the single criterion approach is its failure to differentiate among the size, frequency, duration, and timing of shortfalls. This is a critical omission, as damages associated with power system shortfalls are nonlinear.

Expected unserved energy (EUE) is a preferred addition to incorporate size of shortfalls, especially as the system moves toward energy limitations (p. 15). A first step to better differentiating resource adequacy shortfalls is adding EUE as a resource adequacy criterion. EUE measures the expected (i.e., average) amount of unserved energy per year, averaged across all resource adequacy simulations. A first benefit is that, all other things equal, EUE places a greater emphasis on larger, more disruptive events, a critical consideration in differentiating shortfalls. A second benefit of EUE is that it explicitly measures power system energy limitations—an important consideration as the system becomes more energy-constrained (due to increased storage and load flexibility) and is not just capacity-constrained. In energy-limited systems, the way in which storage or load flexibility is utilized can greatly impact resource adequacy metrics. For example, when a system with short-duration storage capability faces a longer-duration shortfall, it has several options as to how it can deploy its stored energy, each of which yields a different residual shortfall. Operators can choose to use economic criteria to determine dispatch profile, deplete the storage energy as soon as a resource adequacy event starts ("first come, first served"), decrease the duration of the event ("minimize duration"), or decrease the maximum size of the event ("minimize shortfall") (Figure ES-1, p. x). EUE also aligns well with economic metrics, as the value of lost load (VoLL) and other cost metrics are often expressed as $/MWh, facilitating a more straightforward translation between reliability and cost objectives.

FIGURE ES-1

Energy-Limited Resource Scheduling During a Loss-of-Load Event



The figure illustrates how battery storage scheduling can influence resource adequacy metrics. In each case, the total battery storage available is equal to 6 units (blue), and the total unserved energy is equal to 14 units (green). However, decisions of the battery storage scheduling can change LOLE, LOLH, and event characteristics.

Source: Energy Systems Integration Group, adapted from Dent (2019).

No one metric is the solution; a multi-metric framework is needed to consider size, frequency, and duration of shortfalls (p. 33). Given the evolving dynamics in resource adequacy analysis, a changing energy resource mix, and consumer preferences for reliability, adopting a multi-metric criteria approach may be prudent as it provides a more comprehensive assessment of the size, frequency, and duration of shortfalls; explicitly considers tail risks; and can stress-test extreme events that may fall outside historical records.

A multi-metric framework allows planners and regulators to embrace a flexible, multi-dimensional approach that adapts as the risks of the system change. It can also help identify and limit the most impactful risks for a given system and inform stakeholders of the potential size, frequency, duration, and timing of shortfalls.

Specifically Considering Extreme Events

Not all resource adequacy loss-of-load events are the same. Tail risks can have a disproportionate impact on reliability and costs and should be quantified in resource adequacy criteria (p. 22). While traditional resource adequacy studies considered the probability of independent outages occurring at the same time, power system regulators and planners are increasingly concerned about the correlated risk of multiple stressors occurring simultaneously due to underlying weather conditions. These drivers could create tail risks—which are included in resource adequacy analyses, but may occur so seldom that they do not materially influence the average adequacy metrics. However, though they are rare, they are large enough to warrant further analysis and potential investment.

Tail risks can have a disproportionate impact on reliability and costs and should be quantified in resource adequacy criteria.

Such events are akin to a "100-year flood," which is statistically rare but can cause devastating and widespread damage. The cost of damages is often a highly nonlinear function of the size of a power system shortage. For example, a summer evening heat wave causing a shortfall for two to four hours might be far less damaging than a winter event of the same duration or an event of the same duration but much larger in scale.

Limited data are available to confidently determine the probability of extreme events. This reality may require discrete analysis or "stress testing" rather than a statistical measure (p. 29). In some cases, the limited availability of data to confidently determine the probability of extreme events necessitates discrete analysis, or stress testing, rather than relying solely on statistical measures. The inclusion of probabilistic metrics in the planning criteria (like value at risk (VaR) or conditional value at risk (CvaR)), while important, may not be sufficient to ensure system adequacy against rare, high-impact, low-probability events (see Figure ES-2). Deterministic stress-testing evaluates the power system's resilience in specific scenarios, such as a wide-area heat wave, a winter cold snap with limited gas supplies, or a multi-day renewable drought. These events can be explicitly modeled, allowing planners and regulators to understand system risks and prioritize mitigations beyond simply adding new capacity, and offering insights into system vulnerabilities that probabilistic resource adequacy assessments might overlook.

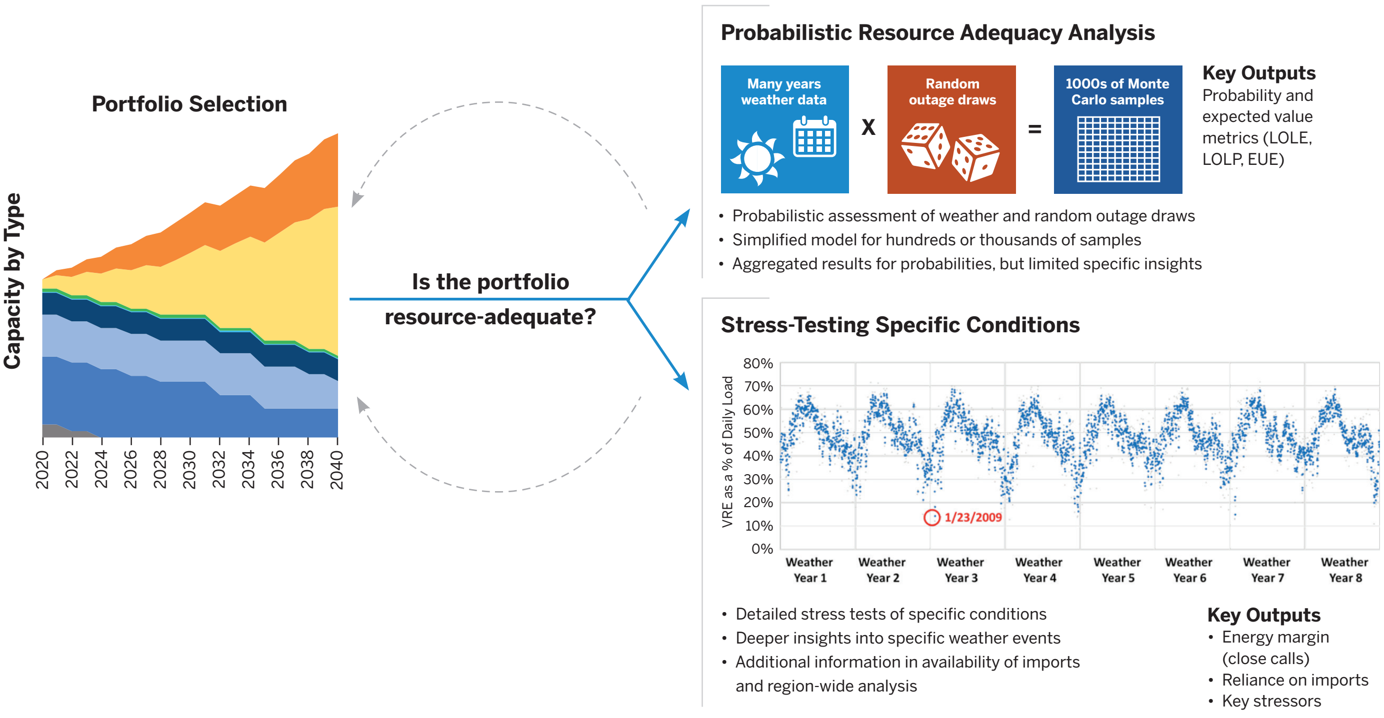
> **Extreme events, such as a wide-area heat wave, a winter cold snap with limited gas supplies, or a multi-day renewable drought, can be explicitly modeled, allowing planners and regulators to understand system risks and prioritize mitigations beyond simply adding new capacity.**

Incorporating Economics

The resource adequacy threshold should be used to establish the appropriate trade-off between reliability and cost. Cost and reliability are intrinsically linked, and this trade-off should be clear (p. 38). Establishing the appropriate level of power system adequacy is not a straightforward task; it requires a collective judgment call involving planners, regulators, customers, and other stakeholders. Resource adequacy in power systems should not be viewed as black and white or a line in the sand. The objective of resource adequacy criteria is to strike the right balance between reliability and cost. Setting the criteria too high can lead

FIGURE ES-2

Combining Probabilistic Resource Adequacy Analysis with Stress Testing



The figure illustrates the two-pronged approach of determining resource adequacy of a resource portfolio, one that includes probabilistic resource adequacy analysis and one that selects challenging time periods for a deterministic stress-testing approach.

Source: GridLab (2022).

Establishing the appropriate level of power system adequacy requires the collective judgment of planners, regulators, customers, and other stakeholders. Resource adequacy criteria need to strike the right balance between reliability and cost.

to prohibitively high investment costs, while setting the threshold too low risks diminished reliability and the potential for significant economic damages (Figure ES-3). It's crucial that this intrinsic link between cost and reliability is transparent and well understood by all involved parties.

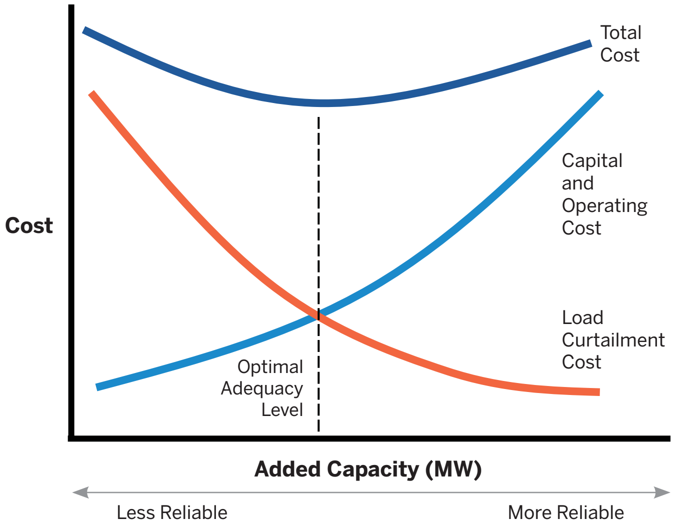
Implementing changes to the reliability standard in the power system requires a broad consensus among various stakeholders. The utilities and grid operators—in consultation with stakeholders—can lead this reform and be the ones to establish the resource adequacy framework, the analytical methods, and metrics used to measure adequacy. They are also responsible for devising specific plans or markets to meet these standards at reasonable costs.

However, the responsibility for determining the level, or minimum threshold, of the resource adequacy criteria ultimately falls on regulators—not the utility or power system planners. They play a crucial role in ensuring that the trade-off between risk and economic factors is appropriately balanced, as this is ultimately a societal and equity decision. This division of responsibilities ensures that while the regulatory bodies establish the criteria, the actual implementation is carried out effectively by those managing the power system.

There are multiple ways to accomplish these goals. To effectively navigate the energy transition, new resource adequacy criteria must encompass, at a minimum, a multi-metric approach, including both LOLE and EUE. Additionally, indicators should capture tail risks, and the framework must be more transparent about providing an economic justification for the chosen reliability level. This comprehensive approach, though challenging, is crucial for ensuring the reliability of our current, and evolving, power systems.

FIGURE ES-3

Optimal Adequacy Level as a Function of Investment Cost and Load Curtailment Damages



The figure illustrates the trade-off between resource adequacy as a function of added capacity (x-axis) and cost (y-axis). As capacity is added to the system, cost (damages) from load curtailment decreases, but the capital and operating costs increase. The optimum level of reliability is where the sum of the two costs, representing total costs, is minimized.

Source: Energy Systems Integration Group.

Introduction and Objectives

Introduction

The transition toward a cleaner and more weather-dependent power system brings with it unprecedented challenges and opportunities for maintaining resource adequacy. Utilities and grid operators face a future that includes fossil-fueled power plant retirements, risk of fuel supply disruptions, load growth, and extreme weather events. These reliability challenges are exacerbated by reliance on variable renewable energy and energy-limited storage resources; changing demand profiles due to electrification of transportation, heating, and industrial processes; and underlying climate change.

As a result of this transition, resource adequacy frameworks are being scrutinized and redefined globally. Resource adequacy is a component of power system reliability that refers to the ability of supply-side, demand-side, and transmission resources to meet the aggregate electricity demand, taking into account scheduled and unscheduled outages of generators and potentially transmission assets. An adequate power supply can accommodate variations in load, fuel supply, and renewable resource production, all of which fluctuate due to weather conditions.

This report builds upon the foundational work laid out in the Energy Systems Integration Group's (ESIG's) 2021 report *Redefining Resource Adequacy for Modern Power Systems*, which highlighted the growing complexities of ensuring a reliable power supply amidst the increasing levels of variable renewable energy and the decline of traditional thermal resources (ESIG, 2021). It provided several recommendations for better representing the effects of a changing resource mix and other challenges in resource adequacy studies. These recommendations, conveyed as six principles, discussed the importance of evaluating chronological operations in resource adequacy studies, incorporating demand-side participation and interregional exchanges, and accounting for uncertainty associated with all resource types. Two of the principles were highly relevant to establishing the resource adequacy criteria and are the focus of this report:

- Quantifying size, frequency, duration, and timing of shortfalls is critical to finding the right resource solutions.
- Resource adequacy criteria should be transparent and should address economic trade-offs.

Building on this original work, ESIG's 2023 report *Ensuring Efficient Reliability: New Design Principles for Capacity Accreditation* emphasized the need to rethink capacity accreditation principles in modern power systems (ESIG, 2023). While the 2021 report focused on system-wide adequacy assessments, this one considered how to assign credit to individual resources for the resource adequacy contributions they provide. It offered recommendations for accreditation methods that are robust, transparent, and reflective of the actual performance and resource adequacy contributions of diverse energy resources.

Table 1 provides an explanation of important resource adequacy terms and definitions. The focus of ESIG's 2021 report was on the first two items: adequacy assessments and studies, and metrics; the 2023 report focused on capacity accreditation; and this report is focused on resource adequacy criteria. All of these are highly interrelated.

While both earlier reports evaluated the changing resource adequacy risks and planning practices created by the energy transition, neither specifically examined potential changes to the resource adequacy criterion used for planning. The resource adequacy criterion (also referred to as the reliability standard or planning criterion) sets the level of supply- and demand-side resources that are required for a given power system to meet reliability objectives. The most common resource adequacy criterion is the loss-of-load expectation (LOLE), which measures the average (also referred to as expected) number of days per year when load is not fully served, whether for a single hour or all 24 hours. Most often, the maximum allowable LOLE is set to 0.1 days/year (or 1 day in 10 years) and is colloquially referred to as the 1-day-in-10 LOLE criterion across much of North America. This criterion is expressed in terms of the expected number of days per year when the load is not fully served for any amount of for any amount of time.

TABLE 1

Resource Adequacy Terms and Definitions

Adequacy assessments and studies	Utilities and system operators perform forward-looking resource adequacy assessments, often probabilistic in nature, to quantify the loss-of-load events across a wide range of system uncertainties. These studies collectively measure the adequacy of the entire power supply.
Resource adequacy metrics	Resource adequacy metrics are outputs from adequacy assessments that quantify risk by measuring the size, frequency, duration, and timing of simulated loss-of-load events. These include metrics like loss-of-load expectation and expected unserved energy.
Capacity accreditation	While resource adequacy studies measure the adequacy of the collective power supply, capacity accreditation measures the ability of individual resources (or classes of resources) to improve overall resource adequacy. Typically this is measured as effective load-carrying capability or using an alternative method.
Resource adequacy criteria	The resource adequacy criterion sets the threshold for an acceptable level of risk. The criterion determines whether a system is deemed adequate and is often converted into a minimum firm (or effective) capacity requirement or planning reserve margin or a capacity market requirement. Resource adequacy criteria are typically expressed by threshold values of corresponding metrics.

Source: Energy Systems Integration Group.

In Europe, the LOLE criterion is defined differently. It expresses unserved load in terms of the average number of shortfall hours per year. In order not to confuse the two criteria, this report refers to the European criterion as loss-of-load hours (LOLH). Despite these nuances, the overall framework is consistent: a single metric that counts only *how often* shortfall days or hours are expected to occur is used as the foundation for determining which investments and resource procurements are made. In both cases, this single metric is used without considering the size, duration, timing, or economic damages associated with the shortfalls.

Limitations of a Single Resource Adequacy Criterion

The resource adequacy criterion is a pivotal standard that influences billions of dollars of investment decisions across both restructured power markets and regulated utilities. It underpins the planning reserve margin, which sets a minimum threshold on the amount of resource needed. In turn, this influences when legacy plants can retire and how many new resources must replace them—whether in vertically integrated planning or in capacity market requirements.

But it now needs to do more—it is not only about capacity shortfalls at any one point in time, but also about energy constraints that arise from variable renewables, increasing battery storage, and limited fuel supplies. Damages from outages increase at a nonlinear rate with increasing duration and size, making it important to differentiate between large or long-lasting outages versus short ones and to consider tail risks associated with high-impact, low-probability events. Our mitigations are also increasingly heterogeneous. Planners now have a portfolio of resources that can meet reliability needs—including wind, solar, storage, and load flexibility—in addition to the gas turbine that has been used historically. The accelerating change in the grid's resource mix and load warrant additional considerations in our long-standing resource adequacy framework.

The resource adequacy criterion now needs to do more—it is not only about capacity shortfalls at any one point in time, but also about energy constraints that arise from variable renewables, increasing battery storage, and limited fuel supplies.

However, while much of the work conducted in the industry over the past few years has focused on how to improve the way we measure resource adequacy risk and contributions of resources, very few studies have sought to answer the questions, "how reliable do we want the system to be?" and "how much money should we invest to improve reliability?" Instead, the industry has continued to use decades-old planning criteria, despite a rapidly evolving grid with a decarbonization landscape that focuses on variable renewable generation, energy-limited storage, and flexible load management.

Moving Beyond a One-Size-Fits-All Criterion

The ageing LOLE criterion has several limitations. First, it is treated as an arbitrary line in the sand rather than articulating resource adequacy as a continuum and a trade-off between costs and reliability. Second, it does not differentiate types of shortfalls, but rather treats tail-end risks associated with longer-duration or larger outages as equal to shorter, less severe outages. Third, it is a static criterion which in many regions has not changed in decades, despite rapid changes to the power system resource mix and electrification of new sectors.

These limitations highlight a need to move beyond a single, one-size-fits-all resource adequacy criterion and augment it with multi-metric criteria.[1] This new approach would leverage a combination of resource adequacy metrics to guide planning and resource procurement and mitigate loss-of-load risk with five central objectives:

- **Limit the likelihood of shortfall events**, which can be accomplished by continuing LOLE-based metrics used in many regions today

---

[1] Throughout this report we use the term **criterion** to refer to a resource adequacy threshold consisting of a **single metric**, and **criteria** to refer to a threshold that includes **multiple metrics**. Regardless of whether the criteria uses multiple metrics or a single metric, the report discusses it as a singular planning construct that a utility or system operator must meet.

- **Address low-probability, high-impact events**, to capture and appropriately mitigate resource adequacy shortfalls that have larger impacts and damages
- **Embrace a flexible, multidimensional approach**, allowing the resource adequacy criteria to adjust as the resource mix and consumer preferences change, and to incorporate a wider array of mitigations including transmission and load flexibility
- **Inform stakeholders of associated risks**, and provide all parties, from system planners to regulators, much more comprehensive information about risks that consumers face
- **Clearly establish the connection between reliability and cost**, establishing the appropriate trade-off between reliability and cost and making the linkage between the two transparent

Why Change Now?

To some it may seem that changing the resource adequacy criterion is unnecessary at this time, and attention would be better spent on continued improvements in resource adequacy studies, simulation methods, and accreditation methods. But many regions are experiencing tightening reserve margins and will require new resources to ensure resource adequacy (NERC, 2023). The risk of waiting to implement change now may lead to costly, immediate capacity deficiencies in the near future. In addition, utility integrated resource planning and capacity market constructs are already undergoing reform, and the stakeholder processes being carried out to update these can be used also to implement changes to the resource adequacy criteria.

Some grid operators are already considering changes. PJM, the Southwest Power Pool (SPP), and the Midcontinent Independent System Operator (MISO), for example, are exploring whether to move from LOLE to expected unserved energy (EUE) as the resource adequacy criterion used for calculating capacity accreditation. The Electric Reliability Council of Texas (ERCOT) and the Northwest Power and Conservation Council (NWPCC) are proposing multi-metric criteria for future resource adequacy planning. Utilities in Colorado are embedding multi-step reliability stress-testing directly in their resource procurement decisions.

However, while these are steps in the right direction, they represent only a small fraction of system operators globally.

This report extends the dialogue on resource adequacy, exploring whether the electric power industry should adopt new resource adequacy criteria as the resource mix becomes increasingly diverse and energy-limited, and as difficult-to-predict tail events may become a greater risk. Under these conditions, the use of a single metric for the resource adequacy criterion will not provide a complete picture of reliability risk. Further, as electricity is projected to be the engine for decarbonization, its value and reliability requirements are of ever-increasing importance.

As the resource mix becomes increasingly diverse and energy-limited, the use of a single metric for the resource adequacy criterion will not provide a complete picture of reliability risk.

Integral to the discussion of adequacy criteria are the inherent trade-offs between reliability and cost. The primary challenge in resource adequacy is to avoid overbuilding the system to eliminate all risk of shortfalls and to ensure that the system's portfolio of resources has the right attributes to ensure reliability. Eliminating all risk in the system is not only prohibitively expensive but also impossible, since planners cannot predict every conceivable challenge the power system will face. As a society, including planners, regulators, policymakers, and ratepayers, we have to decide how much we will pay for reliability and when to accept that there will be times—albeit rare—when the system cannot serve the load. This policy or regulatory decision shapes our resource adequacy criteria, determining the level of supply in which we are willing to invest and recognizing that mitigating all risks on the system may not be worth what that would cost. A resource adequacy criterion does not just measure risk but must also inform actionable investment decisions.

Planners and regulators must weigh different types of investments to improve resource adequacy against a growing set of planning objectives, including costs for ratepayers, environmental objectives, and other options to improve reliability, such as distribution-level outages, transmission stability, or cybersecurity. Different options also come at different costs, which requires decision-makers to evaluate these trade-offs in a consistent manner. Effective multi-metric criteria should enable planners and regulators to delve beyond frequency metrics, like LOLE, to more fully characterize system reliability and better inform these investments.

Objectives of This Report

This report rigorously assesses potential modifications to resource adequacy criteria with the goal of ensuring that the criteria remain relevant and robust in the face of changing generation and demand patterns. Integral to this adjustment are methods that more comprehensively capture the risk to electricity consumers, namely, quantifying the *size, frequency, duration, and timing* of resource adequacy events. Crucially, the report also delves into the trade-off between reliability and cost. The resource adequacy criteria should assist grid planners in selecting an appropriate level of reliability—neither overbuilding the system and paying for diminishing reliability improvements, nor under-investing at the risk of exposing ratepayers to frequent and disruptive outages. Striking the right balance involves a nuanced exploration of how these trade-offs are currently considered and the potential for their more explicit integration into planning processes.

Key adjustments to resource adequacy criteria include more comprehensively capturing the risk to electricity consumers, namely quantifying the size, frequency, duration, and timing of resource adequacy events.

The report offers actionable recommendations for developing and implementing new resource adequacy criteria that will be useful to regulators, planners, industry participants, and other industry stakeholders. It is divided into the following sections:

- An assessment of whether an expanded set of resource adequacy criteria is needed
- The importance of considering size, frequency, duration, and timing of resource adequacy events
- Options for considering tail risks and outlier scenarios
- The importance of stress testing for extreme events
- How to establish multi-metric criteria
- The benefits of including economic considerations when setting adequacy thresholds
- Recommendations for implementation and regular updates to resource adequacy criteria

There are a few things this report does not do. It does not—though it's critically important at this juncture—set specific thresholds for the resource adequacy criteria. Rather, it offers guidelines that planners in different regions can use when discussing options with stakeholders. The report also does not discuss improvements to probabilistic resource adequacy methods or capacity accreditation, and refers readers to previous ESIG work on these topics: *Redefining Resource Adequacy for Modern Power Systems* and *Ensuring Efficient Reliability: New Design Principles for Capacity Accreditation* (ESIG 2021; 2023).

Need for New Resource Adequacy Criteria

Origin and Development of the Probabilistic 1-Day-in-10 Criterion

The use of a resource adequacy criterion, particularly the probabilistic "1-day-in-10 years" loss-of-load expectation prevalent in North America, is a cornerstone in the field of electric power system reliability. While the 1-day-in-10 years, or 0.1 days per year, LOLE criterion is largely a North American construct, regions in Europe and elsewhere use a similar probabilistic LOLE framework. This criterion essentially means that it is acceptable for electric load not to be fully served due to a lack of generation capacity, fuel availability, or transmission for some amount of time on one day every 10 years. This does not mean one day out of a 10-year period (i.e., 2024-2034), nor does it mean one entire day or all 24 hours of a day. Rather, it means that only one shortfall day is allowed out of 10 years of statistical samples of load, weather conditions, and generator outages (Stephen et al., 2022).

It is important to clarify that this is neither a national reliability standard nor a globally recognized norm.[2] Despite its wide use, the exact origins of this specific standard are nebulous, with limited documentation on why and how it was chosen. According to Carden, Wintermantel, and Pfeifenberger (2011), "for decades, the utility industry has been using the 1-in-10 standard as the primary if not sole means for setting target reserve margins and capacity requirements in resource adequacy analyses. While the origination of the 1-in-10 metric is somewhat vague, there are multiple references to it in papers starting with articles by Calabrese from the 1940s."

Loss-of-load expectation as the sole resource adequacy criterion needs to be supplemented—it represents only a single dimension of risk, leaving out size (MW), magnitude (MWh), and duration of resource adequacy events and their potential costs.

Calabrese, an engineer at Consolidated Edison in the 1940s and author of some of the seminal work in probabilistic resource adequacy planning, discussed the level of service reliability that should be aimed for, noting that "a reserve giving a loss-of-load expectancy of 1 day in approximately 50 years, or a probability loss of load of 0.0000548 after voltage reduction, should be satisfactory," and going on to explain that this roughly equates to a 1 day in every 5.6 years probability without voltage reduction (Calabrese, 1950). However, he stressed the importance of tailoring this to specific conditions: "The actual value should be determined after evaluating all local factors involved, making use of information available from past experience."

But even as Calabrese emphasized the need for local factors and judgment, the 1-day-in-10-years criterion has been widely adopted across North America and has remained largely unchanged for decades. According to Carden, Wintermantel, and Pfeifenberger (2011), this enduring acceptance is partly because "customers rarely complain about the level of reliability they receive under

---

[2] Note that 0.1 days/year LOLE does show up in NERC Standard BAL-502-RF-03 (NERC, 2017). In that standard it states as a requirement that the planning coordinator shall conduct a resource adequacy assessment that "calculate[s] a planning reserve margin that will result in the sum of the probabilities for loss of load for the integrated peak hour for all days of each planning year analyzed (per R1.2) being equal to 0.1. (This is comparable to a 'one day in 10 year' criterion)." It does not say that entities must plan to that criterion, but that it must be studied.

the 1-in-10 standard," leading to a lack of scrutiny over its appropriateness.

Both the North American and European metrics are similarly computed and do not characterize the magnitude of shortfalls. In the European Union, a legislative framework was implemented in 2019 that uses an LOLH criterion as a benchmark. The European method intentionally allows disparity in the threshold (reliability standard) across member countries, while establishing a consistent method for resource adequacy assessment. In this framework, each European Union member state can adjust its LOLH thresholds based on individual cost-benefit considerations. This adaptability allows these thresholds to evolve over time, reflecting changing societal and economic landscapes (discussed further in the section "Options Available for Incorporating Economic Principles").

Public perception is a critical factor in the acceptance and understanding of reliability standards, and tolerance of outages presumably will vary significantly based on the size, timing, and duration of outages, and the underlying cause. While the public might tolerate outages due to environmental disasters (hurricanes, tornadoes, floods, earthquakes, etc.), there is generally almost no tolerance for capacity shortfall–related outages due to a perceived lack of planning. This underscores the need not only for setting appropriate reliability standards but also for effectively communicating these standards and their rationale to industry stakeholders and the public at large.

Like many policy decisions, selecting a target level of adequacy requires taking multiple objectives into account, including reliability and affordability. The aim of this report is not to declare an LOLE or LOLH criterion incorrect, but rather to examine deficiencies in using a single metric to summarize system risk and to inquire whether the economic trade-offs involved in setting such a reliability standard are being appropriately and transparently considered.

Resource Adequacy Criteria Used Today

Globally, the approach to establishing minimum levels of resource adequacy varies significantly, reflecting diverse operational, geographical, and economic contexts. Table 2 (p. 8), from EPRI (2022), lists many of the resource adequacy criteria used globally. Five primary metrics are commonly used to set these criteria: loss-of-load expectation (LOLE) in days per year, loss-of-load hours (LOLH) in hours per year, expected unserved energy (EUE) in MWh per year,[3] normalized expected unserved energy (NEUE) measured as the ratio of energy relative to total annual load, and the planning reserve margin usually expressed as a percentage of peak load. While LOLE, LOLH, and EUE are probabilistic metrics, the planning reserve margin is deterministic (but a target planning reserve margin is often derived using probabilistic methods), representing a set level of firm (or effective) capacity above a normal (median) peak load, which is often (but not always) developed based on the results from more detailed probabilistic analysis.

---

[3] EUE is also referred to as expected energy not served (EENS) or expected energy unserved (EEU) depending on region. It can also be normalized as a percentage of load (NEUE).

TABLE 2

Resource Adequacy Criteria Used for Selected Countries and Regions

Country or Region	RA Metrics/Criteria	Entity Calculating RA Metric
North America		
MISO	LOLE ≤ 0.1 days/year	MISO
MRO-Manitoba Hydro	LOLE ≤ 0.1 days/year	Manitoba Public Utilities Board
NPCC-Maritimes	LOLE ≤ 0.1 days/year	Maritimes Sub-areas and NPCC
NPCC-New England	LOLE ≤ 0.1 days/year	ISO-NE and NPCC
NPCC-New York	LOLE ≤ 0.1 days/year	NYSRC and NPCC
NPCC-Ontario	LOLE ≤ 0.1 days/year	IESO and NPCC
NPCC-Québec	LOLE ≤ 0.1 days/year	Hydro-Québec and NPCC
PJM Interconnection	LOLE ≤ 0.1 days/year	PJM Board of Managers
SERC-C	LOLE ≤ 0.1 days/year	Member Utilities
SERC-E	LOLE ≤ 0.1 days/year	Member Utilities
SERC-FP	LOLE ≤ 0.1 days/year	Florida Public Service Commission
SERC-SE	LOLE ≤ 0.1 days/year	Member Utilities
SPP	LOLE ≤ 0.1 days/year	SPP RTO Staff and Stakeholders
TRE-ERCOT	LOLE ≤ 0.1 days/year	ERCOT Board of Directors
WECC-AB	LOLP ≤ 0.02%	WECC
WECC-BC	LOLP ≤ 0.02%	WECC
WECC-NWPP-US & RMRG	LOLE ≤ 0.1 days/year	WECC
WECC-SRSG	LOLP ≤ 0.02%	WECC
WECC-CAMX	PRM ≥ 15% Additional local and flexible RA requirements	CPUC
Hawaii	ERM ≥ 30% (3 islands), 60% (2 islands)	HECO
Europe		
Belgium	LOLH ≤ 3 hours/year	Elia Group
France	LOLH ≤ 3 hours/year	RTE
Great Britain	LOLH ≤ 3 hours/year	National Grid ESO
Ireland and Northern Ireland	LOLH ≤ 8 hours/year (Ireland) LOLH ≤ 4.9 hours/year (Northern Ireland)	EirGrid and SONI
Netherlands	LOLH ≤ 4 hours/year	TenneT
Poland	LOLH ≤ 3 hours/year	PSE
Portugal	LOLH ≤ 5 hours/year	REN
Spain	PRM ≥ 10% (Mainland) LOLE ≤ 1 day in 10 years (Island grids)	REE

(CONTINUED)

TABLE 2

Resource Adequacy Criteria Used for Selected Countries and Regions (CONTINUED)

Country or Region	RA Metrics/Criteria	Entity Calculating RA Metric
Oceania		
Australia-NEM	NEUE ≤ 0.002% per region	AEMO
Australia-NT	NEUE ≤ 0.002%	AEMO
Australia-WEM	PRM ≥ WEM metric NEUE ≤ 0.002%	AEMO
New Zealand	WEM ≥ 14-16% (New Zealand) WEM ≥ 25.5-30% (South Island) WCM ≥ 630-780 MW (North Island)	Transpower
Asia		
India	LOLP ≤ 0.2% NEUE ≤ 0.05%	CEA
Indonesia	PRM (2019-2028) ≥ 30% (National)	Ministry of Energy and Mineral Resources
Japan	PRM (2020-2029) ≥ 8% per region	OCCTO
Laos	PRM (2020-2030) ≥ 15%	Ministry of Energy and Mines
Malaysia	LOLE ≤ 1 days/year	TNB
Philippines	PRM (2017-2040) ≥ 25%	DOE
Singapore	LOLH ≤ 3 hours/year	EMA
Thailand	PRM (2015-2036) ≥ 15%	EGAT
Vietnam	LOLH ≤ 12 hours/year per region	MOIT
Middle East		
Saudi Arabia	PRM (2016) ≥ 8-10%	SEC
Oman	LOLH ≤ 24 hours/year	OPWP
Qatar	PRM (2019) ≥ 6%	KAHRAMAA

Notes: ERM = energy reserve margin; LOLE = loss-of-load expectation; LOLH = loss-of-load hours; LOLP = loss-of-load probability; NEUE = normalized expected unserved energy; PRM = planning reserve margin; WCM = winter capacity margin; WEM = winter energy margin.

Source: EPRI.

LOLE and LOLH are widely accepted as the resource adequacy criterion in North America and Europe, respectively, subject to different interpretations. As noted above, in much of North America, the 0.1 days-per-year LOLE criterion is a widely accepted benchmark. This criterion is expressed in terms of the expected number of days per year when the grid's load is not fully served for any amount of time, whether it is 15 minutes or 24 hours. Conversely, in Europe the more prevalent criterion is LOLH, expressed in terms of the number of hours per year. For example, if resource adequacy shortfalls are relatively short (i.e., lasting 1 to 3 hours), and we plan to one event day every 10 years, this equates to approximately 0.1–0.3 hours per year. This makes the 1-day-in-10 LOLE criterion an order of magnitude more reliable than some other European criteria which are often between 2 and 10 hours per year. For the purposes of this report, both the LOLE criterion and the LOLH criterion are treated as similar, individual-metric criteria.

Australia stands out as one of the few regions utilizing EUE as its resource adequacy criterion. Rather than measure solely the *frequency* of resource adequacy events, this metric *quantifies* the expected amount of energy not served per year, either in terms of megawatt-hours or normalized as a percentage of total load (NEUE).

While tables comparing resource adequacy criteria across different jurisdictions provide a valuable overview of current practices, they overstate the apparent consistency in approaches and miss the fact that many regions are actively considering changes to their criteria. Simply because a metric is widely used does not inherently validate its efficacy or suitability. Each region must independently evaluate the level of risk acceptable for its power system, taking into consideration its unique characteristics and challenges.

Limitations of the Current Use of Single Metrics

The current global landscape for resource adequacy criteria has a defining common feature: the reliance on a single-metric criterion to determine system adequacy and inform investment planning decisions. While this approach offers simplicity, it inherently collapses a vast range of information into a single metric, leaving out crucial details (Felder, 2001). In short, using loss-of-load expectation as the sole resource adequacy criterion represents only a single dimension of risk. The full detail of the system's risk profile cannot be described by a single number; it needs to be supplemented, for the following reasons:

- **A single-metric criterion constitutes a line in the sand, instead of a continuum.** Today, most grid planners and regulators treat the resource adequacy criterion as a bifurcated threshold: the system is either adequate or not. But in reality, power system reliability is a continuum, with various levels of reliability that come at different costs to ratepayers. Too often, the incremental costs associated with reliability—addressing higher or lower levels of risk aversion—are opaque to planners, ratepayers, and regulators. Ensuring transparency in the cost and reliability trade-offs is needed. Directly embedding economic parameters into the criteria is desirable.
- **A single-metric criterion provides inadequate differentiation among the size, frequency, duration, and timing of shortfalls and thus fails to reflect nonlinear damages.** A significant limitation of the single-criterion approach is its failure to differentiate among the size, frequency, duration, and timing of shortfalls. This is a critical omission, as damages associated with power system shortfalls are nonlinear. Longer and larger disruptions lead to disproportionately greater damages, yet the LOLE metric treats all resource adequacy shortfalls equally. This equal weighting does not accurately reflect the real-world impacts of power outages, which vary greatly in severity and consequences. Tail events—those that may occur seldomly but have disproportionately high impacts and costs—require additional focus. A resource adequacy criteria that explicitly differentiates between resource adequacy shortfalls is beneficial.
- **Static criteria are used to represent a dynamic system.** The static nature of the current criterion overlooks the evolving nature of power systems, especially as they become more energy-limited. Historically, there may have been a clear, stable

Power system reliability is a continuum, with various levels of reliability that come at different costs to ratepayers.

Historically, LOLE was a single metric that inherently served as two metrics by both capturing the frequency of shortfalls and serving as a proxy for expected unserved energy, since most resource adequacy shortfalls were similar. However, the mathematical relationship is becoming less stable, leading to the need for additional criteria.

mathematical relationship between loss-of-load expectation and unserved energy, meaning that unserved energy could be approximated by knowing the LOLE. LOLE was a single metric that inherently served as two metrics by capturing the frequency of shortfalls, but also indirectly serving as a proxy for EUE since most resource adequacy shortfalls were similar. However, as the resource mix becomes more diverse and systems become more energy-limited, the mathematical relationship among adequacy measures (e.g., LOLE, LOLH, and EUE) becomes less stable (Fazio and Hua, 2019). Explicitly adding additional criteria (metrics and thresholds) to a resource adequacy standard captures the evolving relationship among these parameters and allows planners to better address customer preferences around reliability.

- **The risk profile is changing as the resource mix evolves.** As the resource mix changes, so does the underlying risk. As the power system becomes more weather-dependent, the types of reliability risks will also shift. Battery storage, for example, may mitigate many short-duration events, but remaining adequacy events become longer and more consequential. In addition, current analytical methods tend to treat weather as an independent uncertainty and, for example, would show resource adequacy shortfalls as being driven by an unlucky coincidence of forced outages of thermal generators occurring during an abnormal high-demand period. Today, however, it is better understood that changes to the resource mix are amplifying *correlated* risks associated with anomalous weather events—for example, a snowy cold snap increasing demand, constraining natural gas supply, shutting down wind turbines due to icing, and covering solar panels with snow. The shifting risk profile necessitates multiple metrics to capture the scope of reliability concerns more comprehensively.

Benefits of Adaptive, Multi-metric Criteria

Today, many resource adequacy planners report a wide array of resource adequacy metrics in their studies and analysis, but investment decisions in capacity markets or integrated resource plans continue to use only a single criterion. Incorporating multi-metric criteria into investment decisions would allow the resource adequacy framework to adapt over time as the system changes. Investment decisions aim to ensure that sufficient resources are acquired so that none of the adequacy metrics exceed their thresholds. Over time, the power supply would become inadequate if one of the metrics were to exceed its threshold, thus making that metric the critical or binding criterion. However, at any given time, one metric may be more influential than the others. As systems evolve, other metrics may increase and even overtake prior binding metrics in influence. All metrics provide supporting information and serve as checks to ensure that resource adequacy remains within predetermined bounds.

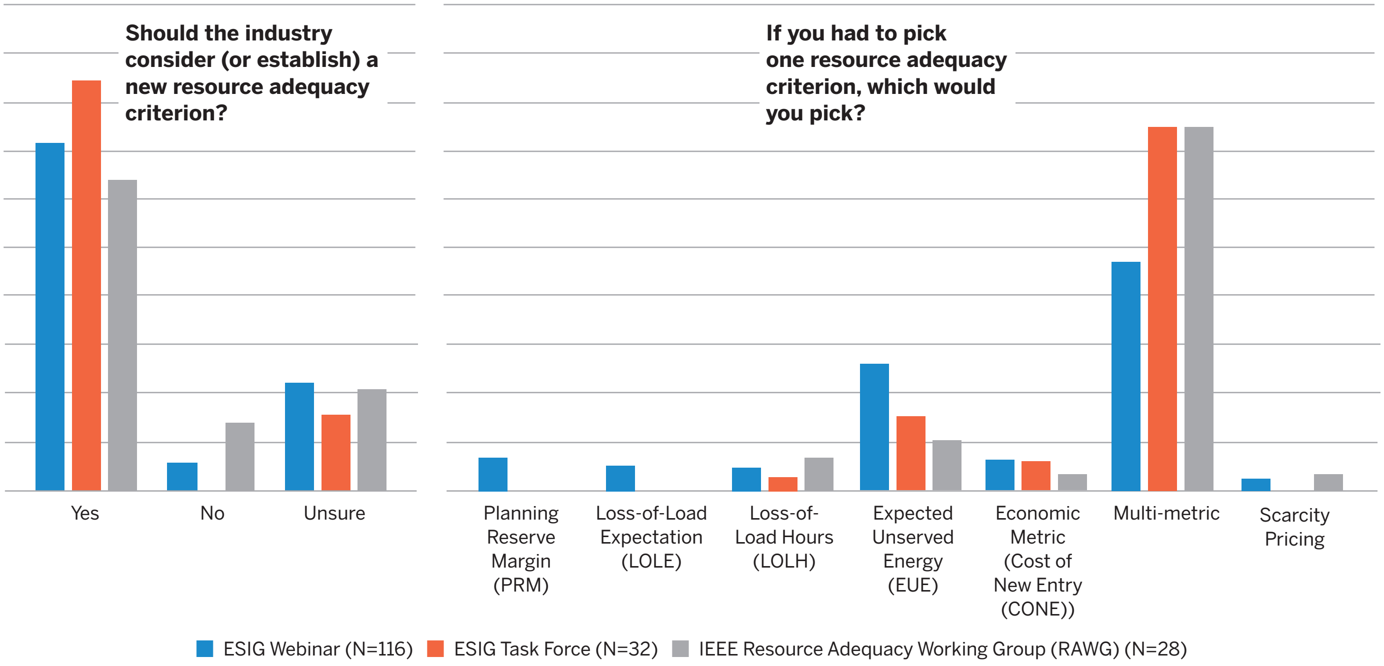
As part of its work, the ESIG task force conducted an informal poll of more than 175 resource adequacy experts and interested stakeholders (see Figure 1, p. 12). The results indicated that the participants believed a new framework for resource adequacy criteria is necessary, with 73% agreeing that a new criteria is justified and 75% preferring a multi-metric approach. Furthermore, the individual metric with the most support was not LOLE or frequency-based metrics commonly used today, but rather EUE.

The survey also solicited feedback on the most important considerations for a new criterion and/or multi-metric criteria. The three features that were ranked the highest included that the criterion or combination:

- Captures tail risks and outlier events
- Measures the magnitude (maximum MW and total MWh) of energy shortfalls and not just the number of times (frequency) that shortfalls occur
- Is easy to understand by utility planners, regional transmission organizations, independent system operators, regulators, and stakeholders

FIGURE 1

Poll Responses Indicating Preference for New, Multi-metric Criteria for Resource Adequacy



Graphs showing the responses to three surveys of resource adequacy experts and stakeholders conducted by ESIG in 2023. Webinar refers to participants in an ESIG webinar on probabilistic methods for resource adequacy, the task force refers to participants in the ESIG Resource Adequacy Task Force, and IEEE RAWG refers to the annual meeting of the IEEE Resource Adequacy Working Group.

Source: Energy Systems Integration Group.

- Explicitly considers the inherent trade-off between cost and reliability

These results indicate a possible growing consensus among experts and stakeholders for the need to evolve beyond the current framework. The challenge for regulators and industry is to determine what the right construct is and how each criterion or criteria should be set. Uniformity in approach should be neither expected nor desired. This report is intended as a guide to provide stakeholders, including grid planners, regulators, and customers, with information needed to develop their own resource adequacy frameworks.

> **The task force poll showed a possible growing consensus among experts and stakeholders for the need to evolve beyond the current framework. The challenge for regulators and industry is to determine what the right construct is and how each criterion or criteria should be set.**

Progress on New Criteria

Consideration of new frameworks for resource adequacy criteria is already underway in various jurisdictions including developments in the European Union; consideration of a potential transition to EUE in PJM, SPP, and MISO; multi-metric criteria being proposed in ERCOT and NWPCC; and a multi-part stress-testing approach being considered by utilities in Colorado. This section provides a high-level overview of the progress being made in different regions, indicating a move toward more nuanced and multi-dimensional resource adequacy frameworks. However, this is a rapidly changing area, and various regions are in the process of revising or changing their resource adequacy constructs.

As a result, these observations may quickly become out of date.

European Union: New Standards for Resource Adequacy Criteria

In 2019, the European Union adopted a legislative framework on processes and standards for assessing resource adequacy. These stipulations mandate each member state to develop an LOLH criterion, require reporting of EUE, and standardize the process for setting the criteria by considering the cost of new entry (CONE) and value of lost load (VoLL), as per Article 23(6) of Regulation (EU) 2019/943 (EU, 2019). Moreover, the regulation established a regular pan-European Resource Adequacy Assessment (ERAA) process, which annually monitors the development of resource adequacy in the bidding zones of the European internal electricity market for the next decade. The ERAA may be supplemented by national resource adequacy assessments (NRAAs) by the European Union member states. In case of a resource adequacy concern identified by the ERAA or a national resource adequacy assessment, a member state would be entitled to introduce a capacity mechanism to safeguard adequacy.

PJM, SPP, and MISO: Considering a Transition to Expected Unserved Energy

PJM is currently undergoing a capacity market reform, part of which considers shifting to EUE as the primary resource adequacy metric used in resource accreditation (PJM Interconnection, 2023b). In this process the 1-day-in-10-years criterion will continue to be used but will be translated into an equivalent level of EUE in megawatt-hours. Similar proposals to consider EUE as a planning criterion or for resource accreditation is also being considered by SPP and MISO (SPP, 2024). However, in each of these jurisdictions, consideration of a new criterion is in its infancy, and final proposals have not been decided. For example, MISO's current goal is to "evaluate and deliver a roadmap incorporating the Expected Unserved Energy (EUE) reliability metric" (MISO, 2023).

ERCOT and the Public Utility Commission of Texas: A Three-Part Framework

ERCOT and the Texas Public Utility Commission are developing a new reliability standard for the ERCOT market (PUCT, 2023). The new metrics limit the frequency of any shortfall event by setting a threshold for loss-of-load events (LOLEv) and limit the size and length of the 97.5th percentile (VaR97.5) tail-end event by setting thresholds for its duration (hours) and energy (MWh) and peak (MW) losses. The currently proposed provisional limits will be reviewed and potentially amended following further analysis and stakeholder feedback.

Northwest Power and Conservation Council: A Four-Part Criteria

The Northwest Power and Conservation Council (NWPCC) has developed a new four-part resource adequacy standard (NWPCC, 2024). The new metrics include an LOLEv metric to set a frequency threshold and three additional 97.5th percentile tail-end metrics (value at risk or VaR), to limit shortfall duration (hours), single-hour peak (MW), and annual energy (MWh). These provisional limits will be reviewed and potentially amended following further analysis and stakeholder feedback.

Colorado Utilities: Stress Testing Included in the Resource Adequacy Criteria

Tri-State Generation and Transmission Association (Tri-State) and Public Service of Colorado have incorporated multi-staged resource adequacy criteria in their latest Energy Resource Plans that include stress testing of portfolios. For instance, Tri-State's framework includes Level 1 metrics (i.e., 1-day-in-10-years LOLE) and Level 2 metrics (evaluating adequacy explicitly during extreme weather events, where systems must meet multiple LOLH and EUE criteria) (Tri-State Generation and Transmission Association, 2023).

> **Developments across various jurisdictions reflect a growing awareness of the changing nature of resource adequacy in light of new resource mixes, the need to balance economic considerations, and the desire to avoid significant tail-end events. These innovative approaches could serve as models for other regions.**

These developments across various jurisdictions reflect a growing awareness of the changing nature of resource adequacy in light of new resource mixes, the need to balance economic considerations, and the desire to avoid significant tail-end events. As the electricity sector continues to evolve, these innovative approaches to resource adequacy criteria could serve as models for other regions grappling with similar challenges.

However, while these examples are steps in the right direction, they represent only a small fraction of system operators globally, are only in early stages of development in most places, and are not yet formalized in procedural or regulatory processes. They are also ad hoc proposals, often tied to a single regulatory requirement, and do not consider other changes to the resource adequacy framework (like accreditation reform) or proposals in other jurisdictions.

Capturing Size, Frequency, Duration, and Timing

While LOLE offers insights into the frequency of shortage conditions, it falls short in providing a comprehensive picture of shortfall characteristics that are crucial for regulating reliability and informed investment decision-making. Given the integration of new energy-limited resources like variable renewables, load flexibility, and energy storage, the one-size-fits-all approach of this traditional metric is increasingly insufficient. The increasing diversity of resource mixes changes the nature of risk—the frequency, duration, size, and timing of shortfalls is evolving. Each new resource type contributes differently toward resource adequacy and risk mitigation. Planners must assess each new resource's firm contribution to select the types and sizes of resources that, together, best fit the resource adequacy needs of the system. Effective criteria should enable planners and regulators to delve beyond frequency metrics to answer other critical questions related to system reliability and investment, including:

- **Expected annual shortfalls:** Measuring total energy not served and the number of hours, days, or events with a shortfall remains critical.
- **Average event characteristics:** Knowing the typical characteristics of shortfall events enables system design that can handle common disruptions effectively.
- **Credible worst-case scenarios:** Identifying tail events, or extreme cases, which are rare but have a high impact, is crucial for developing resilience against extreme situations.
- **Range of possible events:** Assessing the variance of potential events, including their size and duration, provides a more nuanced understanding of system vulnerabilities.

Expected unserved energy is a preferred addition to incorporate size of shortfalls, especially as the electric power system moves toward energy limitations.

- **Cost of mitigation:** Estimating the economic implications of various mitigation strategies is essential for allowing planners and regulators to balance reliability with affordability.

Multi-dimensional resource adequacy criteria that provide additional information to characterize a full spectrum of risk—including consideration of size, frequency, duration, and timing of shortfalls—can better inform the economic-risk trade-off.

Differentiating Among Types of Resource Adequacy Shortfalls

The traditional criterion, which relies on single metrics like LOLE or LOLH, presents two significant limitations: it only offers a single, average index to summarize a spectrum of shortfall events, and it fails to provide any information regarding the severity of events. These deficiencies are critical, as not all shortfall events have the same impact; a multi-day shortfall event in winter, for instance, poses far greater challenges, costs, and threats to human life than a short-lived event in summer.

As noted above, historically, shortfall events were treated only as binary outcomes in metrics like LOLEv and LOLE—either there is load loss or there is not. But this

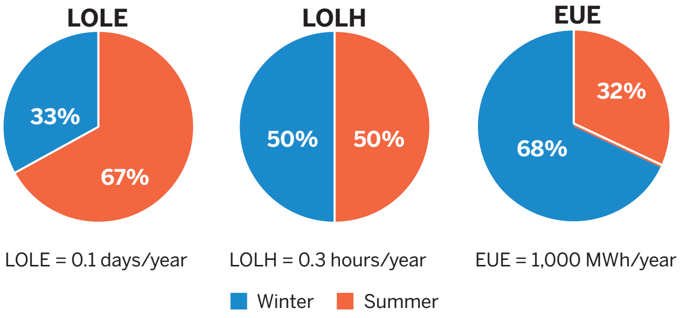
perspective neglects the magnitude of shortfalls, which is crucial, as larger shortfalls have increasingly more severe implications. EUE partially addresses this issue by providing the expected amount of unserved energy (MWh of curtailed demand) over a specified time period, regardless of the number of shortfall events during that period. However, single-dimensional metrics, which are typically represented by aggregated and averaged values like LOLE and EUE, inevitably lead to information loss, with each unable to differentiate between various combinations of event characteristics (Stenclik et al., 2022).

Furthermore, different metrics provide information about different aspects of risk, especially on a seasonal basis. Winter events, which are often characterized by sustained high loads in regions with electric heating and potential fuel supply constraints, can look very different from summer events. Figure 2 shows model results from a resource adequacy analysis in PJM, comparing EUE, LOLH, and LOLE, separated by season (PJM Interconnection, 2023a). This shows how a single metric, used in isolation, can miss important nuances in a region's risk profile. In this case, using LOLE could bias results to favor mitigations available in summer periods, while EUE will do the opposite. In other words, from Figure 2 it can be inferred that there are likely going to be more shortfall events in summer (67% of LOLE days occur in the summer), but these events are shorter than winter events (50% of loss-of-load hours occur in only 33% of loss-of-load days). And they are much larger, as just 33% of the loss-of-load days constitute 68% of the unserved energy.

We can see the importance of multifaceted criteria in recent as well as historical resource adequacy events in regions like California (summer 2020 heat wave), Texas (February 2021 Winter Storm Uri), and the Southeast (December 2022 Winter Storm Elliott). For instance, while these events may be characterized similarly when they are summarized in the LOLE metric, they differ

FIGURE 2

A Comparison of Resource Adequacy Metrics by Season, PJM



The charts show the seasonal results of a single PJM resource adequacy simulation, indicating how different metrics can summarize seasonal risk differently. On the left is LOLE, in days per year, segmented by season and showing risk concentrated in summer. In the center is LOLH, in hours per year, being equally distributed between summer and winter seasons. On the right is EUE, in MWh per year, showing a concentration of unserved energy in the winter season. (Model results are preliminary, subject to change, and not intended to indicate a specific future planning year.)

Notes: EUE = expected unserved energy; LOLE = loss-of-load expectation; LOLH = loss-of-load hours.

Source: PJM (2023b).

TABLE 3

Comparison of Resource Adequacy Metrics in California (2020), Texas (2021), and the Southeast (2022)

Event Characteristic	Metric Affected	California August 2020	Texas February 2021	Southeast December 2022
Number of days	LOLE	2 days	4 days	2 days
Number of events	LOLEv	2 events	1 event	2 events
Number of hours	LOLH	6 hours	71 hours	7 hours
Unserved energy	EUE	2,700 MWh	990,000 MWh	40,000 MWh
Maximum shortfall		1,072 MW	20,000+ MW	5,400 MW

The table shows the measures of unserved energy in three recent loss-of-load events in the United States between 2020 and 2022 and how they would influence resource adequacy metrics.

Notes: Southeast event includes TVA (7 hours, DEC (3 hours), DEP (2 hours), LG&E/KU (4 hours), DESC (9 minutes), and Santee Cooper (17 minutes).

Notes: EUE = expected unserved energy; LOLE = loss-of-load expectation; LOLEv = loss-of-load events; LOLH = loss-of-load hours.

Source: Energy Systems Integration Group.

dramatically in terms of unserved energy, especially for the Texas event, whose unserved energy was 23 times higher than the Southeast event and 366 times higher than the California event (Table 3).

While EUE provides a measure of the overall size of load loss, it still collapses the system's risk profile into a single number. For example, it does not provide any information about individual shortfall event size: a system with many small events can have the same EUE as a system with a few large events. A system with frequent, small shortfalls—causing frequent disruption to customers—may lead to customer fatigue and political pressure to change system planning or market rules. To effectively illustrate these differences and inform system planning, emergency actions, and mitigation strategies, a suite of criteria may be preferred. This suite should encompass metrics that characterize the size, frequency, duration, and timing of shortfall events, moving beyond single-point assessments to include distributions and visualizations.

An example of this differentiation can be seen in Figure 3 (p. 18), which shows the event distributions characterized by event duration in hours (x-axis) and maximum event size (y-axis) in a resource adequacy study of ERCOT (EPRI, 2023b). In both cases, the LOLE is exactly the same (0.1 day/year). However, the results show that in a future, higher-renewable system with increased reliance on renewables and storage, the risk shifts to winter months and the probability of large events increases.

While the adoption of multiple metrics can enhance the understanding of potential shortfalls, the challenge lies in translating this suite of information into actionable criteria for informing regulation, planning, and investment decisions.

Expected Unserved Energy as a Criterion

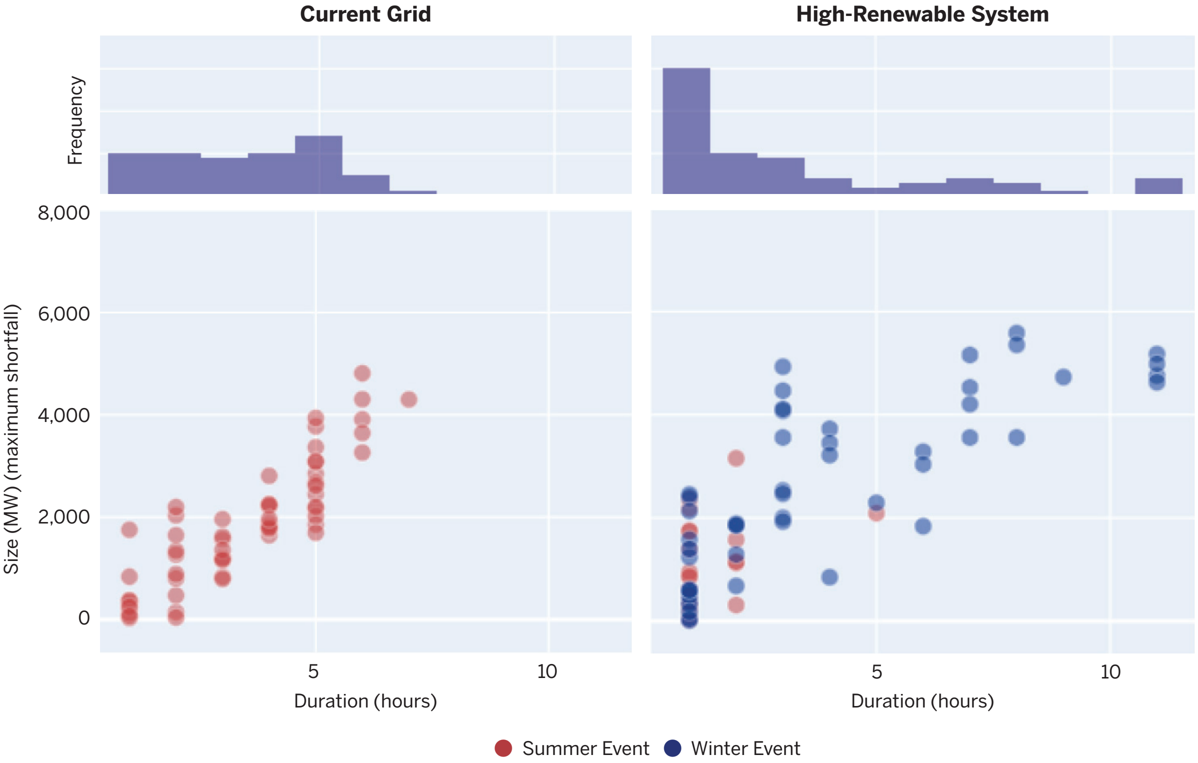
A first step toward better differentiating resource adequacy shortfalls is adding EUE as a resource adequacy criterion.[4] EUE measures the expected (i.e., average) amount of unserved energy per year, probability-weighted across all resource adequacy simulations. For a simplified example, if a resource adequacy study includes 10 sample years, but only one experiences loss of load (i.e., 1-day-in-10-years LOLE) of 3,000 MWh, then the EUE would be 3,000 MWh / 10 years, or 300 MWh/year.

---

[4] In Europe and other countries EUE is also referred to as expected energy not served (EENS).

FIGURE 3

Differentiation of Shortfall Events in ERCOT Between a Current System and High-Renewable System



The scatter plots summarize the loss-of-load events calculated in a resource adequacy study of the ERCOT system. The two plots compare a representation of the current grid (left) and a higher-renewable future grid (right). Each data point represents a loss-of-load event collected across 600 samples and summarized by duration (x-axis) and maximum size of the shortfall (y-axis), and colored by season. The top portion of the chart shows a histogram of duration frequency.

Source: EPRI (2023b).

Oftentimes the EUE metric is normalized (referred to as NEUE) and reported as a percentage of annual load to allow the metric to be compared across years or between different systems.

Today, the EUE metric is used as a single criterion in Australia (AEMC, 2023), but it is often reported globally as a secondary metric. As mentioned above, in the European Union, for example, EUE (referred to as expected energy not served (EENS)) must be reported by all member states alongside the LOLH criterion reliability standard (EU, 2019). However, European Union member states typically define the reliability standard in terms of LOLH. While LOLH remains the resource adequacy criterion that determines when new resources are required for resource adequacy, the EUE metric must be reported to provide additional insight

TABLE 4

Benefits and Limitations of Expected Unserved Energy as a Resource Adequacy Criterion

Benefits of EUE as an RA Criterion	Limitations of EUE as an RA Criterion
Incorporates size of shortfall events	Does not explicitly capture the *frequency* of shortfalls
Places higher weight on large, disruptive tail events	Can overlook frequent but small events that may be inconvenient to customers
Is easier to translate to an economic value by assigning a value of lost load (VoLL)	Normalized EUE (NEUE) relative to system load can be difficult to interpret
Better accounts for energy limitations of storage and load flexibility resources	We have limited experience setting EUE-based reliability criteria, and they are more difficult to understand

Source: Energy Systems Integration Group.

and comparison across member countries. In the United States, grid operators like PJM and SPP are considering adopting EUE as the resource adequacy criterion used to accredit resources and to establish the capacity that load-serving entities (e.g., utilities) must build or purchase (PJM Interconnection, 2023a; SPP, 2024).

While EUE has its benefits, it also presents limitations, outlined in Table 4.

A first benefit is that, all other things equal, EUE places a greater emphasis on larger, more disruptive events, a critical consideration in differentiating shortfalls. In future energy systems where wind, solar, and battery energy storage play a more significant role, resources like battery storage can effectively reduce the frequency of short-duration events. For example, today's resource adequacy risk is often characterized by a few select hours during the period of peak demand—a challenge that can be effectively mitigated by four-hour battery storage. However, those resources are less effective at mitigating longer-duration events, like winter cold snaps that elevate demand for sustained periods of time. As a result, energy-limited resources may mitigate all of the shorter-duration events (thus significantly reducing LOLE or the frequency of events), but the events that remain are disproportionately large (thus keeping the EUE high).

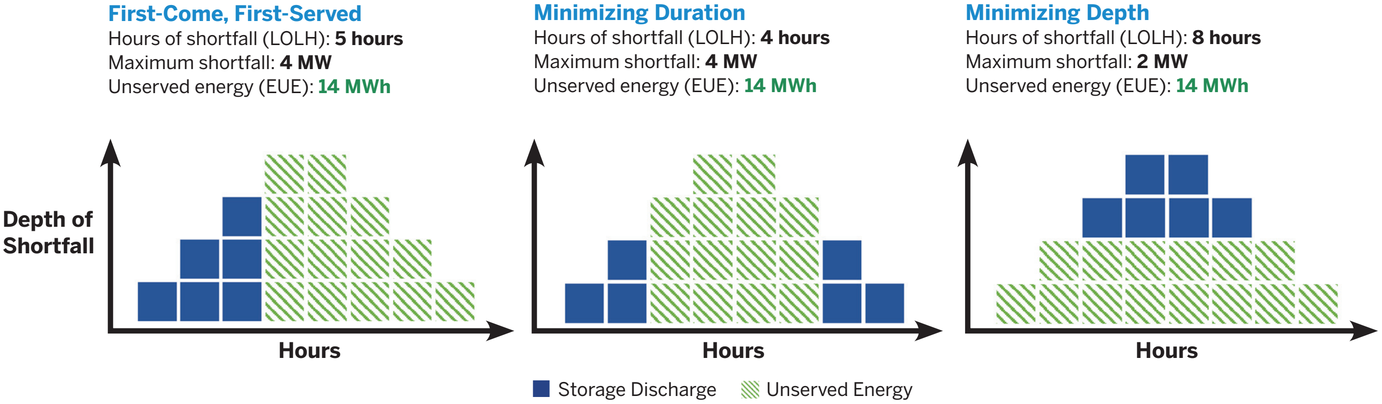
The LOLE value is not an accurate indicator for EUE because the relationship between these metrics changes as resource mixes and load profiles evolve.

Put another way, the LOLE value is not an accurate indicator for EUE because the relationship between these metrics changes as resource mixes and load profiles evolve. For energy-limited systems, for example, loss-of-load events can remain relatively infrequent, but when they do occur, their magnitudes tend to be larger. Systems with similar LOLE values can have vastly differently sized shortfall events. Thus, to provide a more comprehensive assessment of resource adequacy, both metrics are needed.

A second benefit of EUE is that it explicitly measures power system energy limitations—an important consideration as the system becomes more energy-constrained (due to increased storage and load flexibility) and is not just capacity-constrained. In energy-limited systems the way in which storage or load flexibility is utilized can greatly impact resource adequacy metrics. For example, when a system with short-duration storage capability faces a longer-duration shortfall, it has several options as to how it can deploy its stored energy, each of which yields a different residual shortfall (Gonzato, Bruninx, and Delaure, 2023; Stephen et al., 2022). Operators can choose to:

FIGURE 4

Energy-Limited Resource Scheduling During a Loss-of-Load Event



The figure illustrates how battery storage scheduling can influence resource adequacy metrics. In each case, the total battery storage available is equal to 6 units (blue), and the total unserved energy is equal to 14 units (green). However, decisions of the battery storage scheduling can change LOLE, LOLH, and event characteristics.

Source: Energy Systems Integration Group, adapted from Dent (2019).

- Use economic criteria to determine dispatch profile (an "economic" option),
- Deplete the storage energy as soon as a resource adequacy event starts ("first come, first served"),
- Decrease the duration of the event ("minimize duration"), or
- Decrease the maximum size of the event ("minimize shortfall").

An illustration of these trade-offs is provided in Figure 4.

To illustrate the impact of these scheduling decisions, consider the ERCOT case study presented in Figure 3 above (p. 18). In this example, a future system with increased variable renewables and energy storage was evaluated testing different methods of scheduling storage resources. The results, provided in Table 5 (p. 21), show that the resulting non-EUE resource adequacy metrics can vary significantly for the exact same system when only changing the way battery storage is scheduled. Underpinning the difference in storage dispatch, the observed EUE is the same because the available energy from storage is identical going into each event; the difference is in which hours it is discharged. This stability in EUE results under different dispatch conditions—while other metrics vary considerably—underscores EUE's robustness in energy-limited systems, while also highlighting the importance of considering EUE alongside these additional metrics.

EUE also aligns well with economic metrics, as the VoLL and other cost metrics are often expressed as $/MWh, facilitating a more straightforward translation between reliability and cost objectives. Additionally, EUE can be measured by season, month, or hour of the day, which allows for a more accurate assessment of cost if a seasonal VoLL is determined.

TABLE 5

Comparison of Resource Adequacy Metrics Based on Storage Scheduling Objectives

	First Come, First Served	Minimize Duration (hrs)	Minimize Depth (MW)
LOLE (days/year)	0.08	0.10	0.11
LOLH (hours/year)	0.23	0.39	0.56
EUE (MWh/year)	725	725	725
Average depth (GW)	3.9	2.2	1.4
Maximum depth (GW)	9.0	6.3	4.3
Average duration (hours)	2.8	2.8	5.8

The table summarizes resource adequacy metrics from a study on the ERCOT system. In each column, the same simulation was run, changing only the way battery storage was scheduled by the model, showing how LOLE, LOLH, and event characteristics can change based on the scheduling objectives, but EUE remains constant. The highlighted boxes show the metric being optimized by the simulation.

Notes: ERCOT = Electric Reliability Council of Texas; EUE = expected unserved energy; LOLE = loss-of-load expectation; LOLH = loss-of-load hours.

Source: EPRI (2023b).

However, EUE alone may not be sufficient. Frequency-based metrics like LOLE still have their use. Understanding the frequency of shortfalls is important, because frequent smaller events may be less tolerable to consumers and therefore more politically sensitive than larger events, even if they cause less economic damage. Furthermore, unserved energy, measured in MWh/year (EUE)—or as a percentage of annual load (NEUE)—is less relatable to stakeholders than metrics that measure the expected number of hours or days in a year that could have shortfalls.

The Australian Energy Market Operator (AEMO) uses EUE as the standard for the National Electricity Market (a grid covering around 80% of Australian demand), and a multi-metric standard incorporating EUE and a reserve standard for the Western Energy Market. The EUE standard in the National Electricity Market is well regarded; however, there are concerns that it may not adequately capture tail risk, and a process is underway to consider introducing a multi-metric standard incorporating tail measures such as conditional value at risk (CVaR). In the interim, a tighter reliability standard has been adopted to incorporate tail risk using an EUE measure (AEMC, 2023).

Considering Tail Risk

What Are Tail Risks, and Why Do They Warrant Particular Attention?

Not all resource adequacy loss-of-load events have the same impacts. Tail risks can have a disproportionate impact on reliability and costs and should be quantified in resource adequacy criteria.

As the resource mix changes and the power system becomes increasingly weather-dependent, the potential for large, correlated supply disruptions increases. While traditional resource adequacy studies consider the probability of independent outages occurring at the same time, power system regulators and planners are increasingly concerned about the correlated risk of multiple stressors occurring simultaneously due to underlying weather conditions. For example, an extreme cold snap could bring weather conditions that elevate demand for a sustained period (further amplified due to electric vehicles and electrification of space heating), increased generator outages, decreased natural gas supply, icing on wind turbines, and snow cover on PV panels. These drivers could create tail risks—which are included in resource adequacy analyses but may occur so seldom they do not materially influence the average adequacy metrics. However, though they are rare, they are large enough to warrant further analysis and potential investment.

It is crucial to incorporate a metric or process that specifically addresses tail risks, with their disproportionately large impacts on the power system. A well-planned system might have, for instance, a 1-day-in-10-years threshold limit on the frequency of not serving some load for a single day. However, within this framework, very large, long-duration events might occur once every 30, 40, or even 100 years. These outlier events, characterized by their significant maximum peak shortfall (MW), energy shortfall (MWh), or duration (hours), can lead to exceedingly large economic damages and loss of life.

Such events are akin to a "100-year flood," which is statistically rare but can cause devastating and widespread damage. Insurance companies, emergency services, and individuals may place greater emphasis on preparing for these large-scale natural disasters than they do for minor floods, even if the "expected" total cost of minor floods (damages times the probability) is higher than the 1-in-100 year event. Individuals purchase health, auto, and home insurance to limit the potential damages of extreme events that could jeopardize their financial stability. Mitigating power system risk is similar, but today's single metric focused on the frequency of shortfall events does not capture the likelihood, magnitude, or economic impacts of tail events.

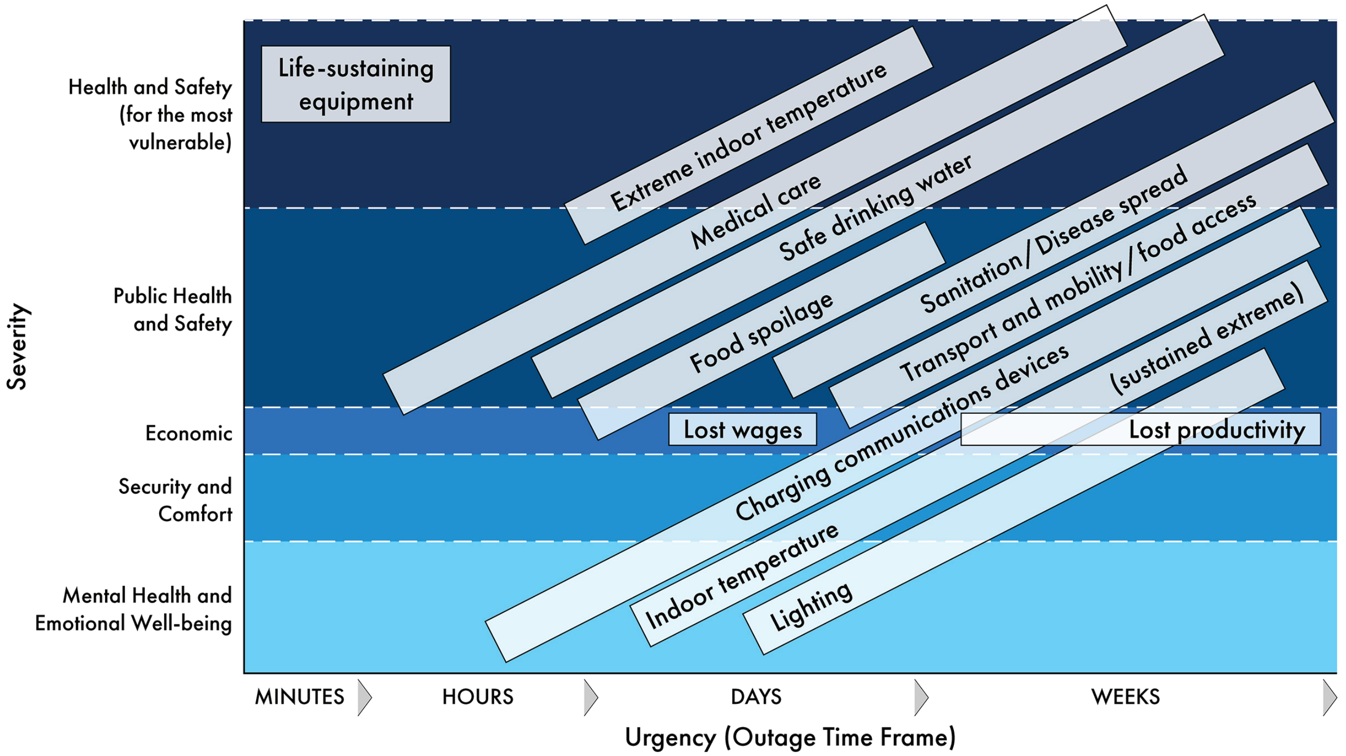
A real-world example of such a tail event is Winter Storm Uri, which significantly impacted Texas, creating a situation well beyond the average 1-day-in-10-years shortfall. This event lasted for multiple days of continuous load loss at a level that exceeded 30-40% of total electricity demand. It was one of the largest disruptions of the North American power system on record. Even if an event like Winter Storm Uri were only to occur once every 50 years without any other resource adequacy shortfalls during that period, it is obvious that this outcome would be unacceptable from a risk mitigation perspective given the magnitude of economic damages and loss of life.

EUE, used in isolation, may still not drive the necessary investment to mitigate these rare, extreme events. In particular, it still collapses risk into a single metric. This is because traditional metrics like LOLE, LOLH, and EUE are expected values that average across all events and will smooth out the impact of such outlier events, thus making their impacts on the average resource adequacy metrics indistinguishable.

Another important consideration for metrics that capture tail risk is their use in system operators' and regulators' efforts to assess economic damages associated with these types of scarcity events. The cost of damages is often a highly nonlinear function of the size of the event. For example, a summer evening heat wave causing a shortfall for two to four hours might be far less damaging than a winter event of the same duration or an event of the same duration but much larger in scale. An example of the severity versus duration of shortfall events is provided in Figure 5.

FIGURE 5

Customer Damage as a Function of Outage Duration



The figure illustrates the customer damages associated with loss-of-load events as a function of outage time (x-axis) and severity of damages (y-axis). The upward slope shows how increasing duration of outages lead to increased damages. The segmentation of customer damages show differences in the value of lost load based on end use.

Source: EPRI (2023a).

As a result, incorporating tail risk into the resource adequacy criteria is essential for a comprehensive assessment of system reliability. It involves acknowledging and taking precautions for the unlikely but potentially catastrophic events that could significantly disrupt power systems. This approach ensures careful evaluation and assessment of the trade-offs associated with infrequent but large disruptions.

Changing Tail Risks with a Changing Resource Mix

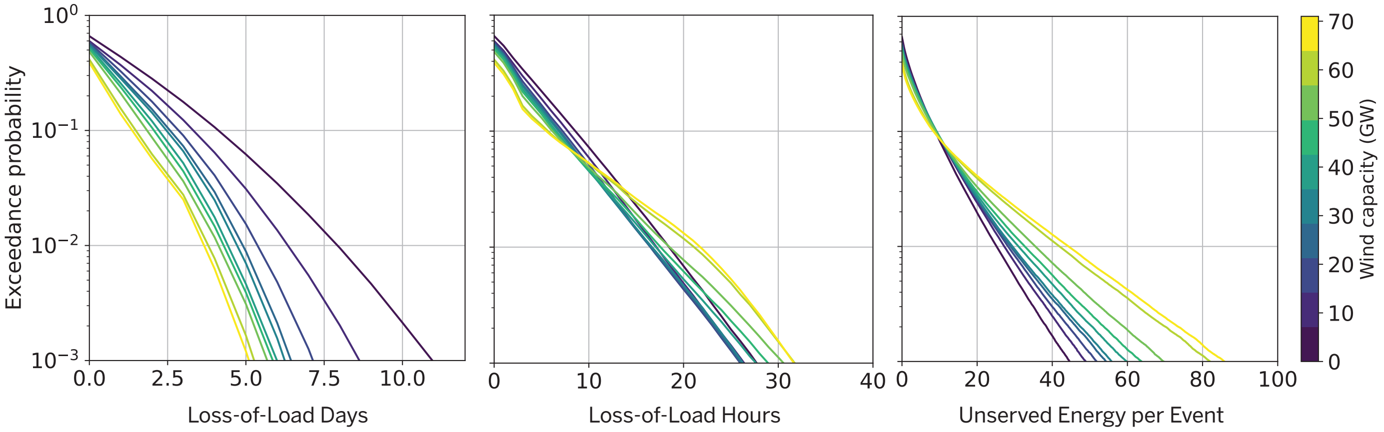
As the power system's resource mix changes, assessing tail risks becomes increasingly important to consider. Most notably, as the system becomes more weather-dependent due to increased variable renewables and end-use electrification, the resource adequacy risk will start to concentrate into a limited number of specific weather events. This consolidation of the worst events can get washed out, or averaged away, in analyses based on historical records.

Tail risks are evident in recent research findings that show that risk is often concentrated over a small number of days, generally associated with extreme or anomalous weather (EPRI, 2023b). Even when using a long meteorological history, there is high uncertainty whether the sampling is representative. Estimates of adequacy metrics are driven by a very limited number of historical years. Moreover, much existing analysis implicitly assumes that one can meaningfully specify a single value for risk metrics, which is not conditional on particular weather patterns. But this assumption is likely not accurate for systems with increased weather sensitivity where the statistical characterization of weather is tentative, especially in the face of climate change. Not having a long meteorological history, and a changing climate, clearly has implications for how reliability standards are set.

The changing tail risks associated with a changing resource mix is observed when analyzing forward-looking scenarios for power systems like that in Great Britain. Research by Dent et al. (2023) demonstrates that even while maintaining a constant EUE, the risk profile can vary significantly (Figure 6). In scenarios where variable renewables play a more dominant role (lighter colored, yellow lines across all charts), the overall risk profile (measured in terms of constant EUE) may comprise fewer (left panel), yet more severe (center and right

FIGURE 6

Risk Profile of the Great Britain System with Fixed EUE, but Varying Installed Wind Capacity



The figure illustrates the difference in underlying resource adequacy metrics on a future Great Britain system with different amounts of installed wind capacity (colored lines) and a constant EUE across all cases. The panels show a probability distribution of loss-of-load days (left), loss-of-load hours (middle), and unserved energy within each day with lost load (right). The results show that in a high wind scenario, the number of days with lost load is lower, but when these events do occur, they tend to be longer and more severe (higher unserved energy).

Source: Dent et al. (2023).

panels), shortfall days. This finding indicates a shift in the nature of risk from more frequent, smaller events to less frequent, but larger and potentially more disruptive events.

Such insights underscore the necessity of re-evaluating how tail risks are considered in the context of resource adequacy criteria. Since it is highly impractical or prohibitively expensive to eliminate all risks and achieve a 100% reliable system, evaluating the potential tail risks—and the underlying weather drivers—is essential. By understanding these risks, planners and regulators can develop resource adequacy thresholds, planning practices, market rules, and other administrative mechanisms that prompt the necessary investments to limit the likelihood and impact of these extreme outlier events.

Options for Tail Risk Criteria

Visualizing Distributions and Modeling Results

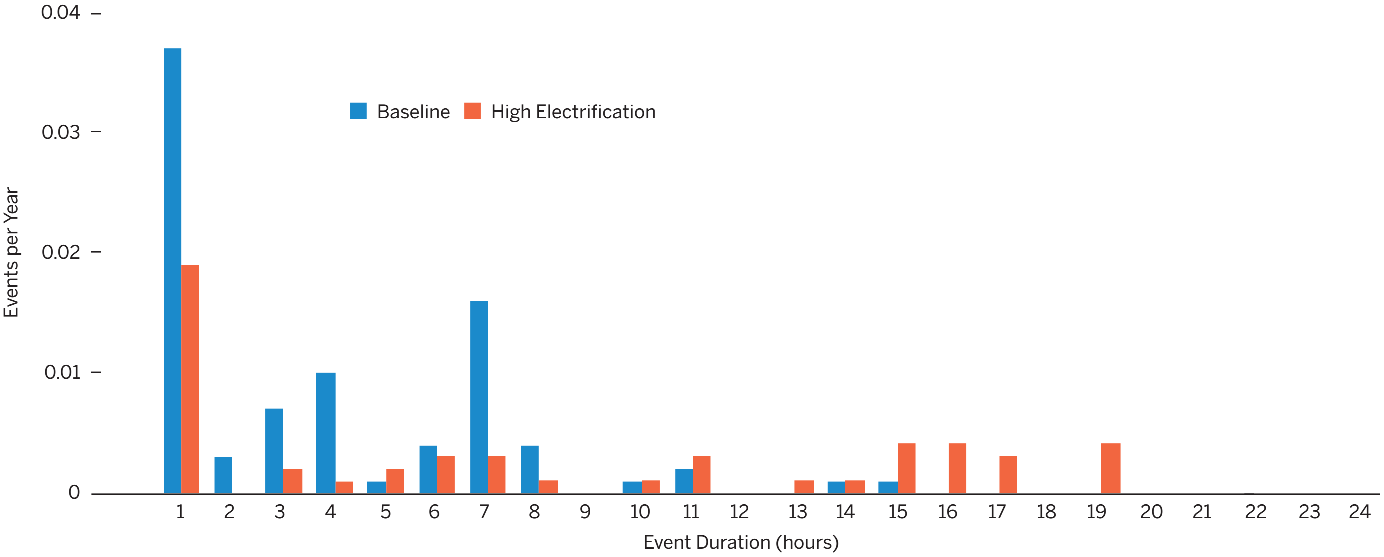
There are multiple options for incorporating tail risks into resource adequacy studies and the resource adequacy criteria. First and foremost, resource adequacy studies are increasingly reporting the full distribution of shortfalls rather than just average or expected values. Even without explicitly embedding tail-risk metrics into a resource adequacy criteria, detailed reporting can convey valuable information about risk to decision-makers. This involves presenting distributions that show values for maximum capacity shortfall size (MW), maximum energy shortfall (MWh), or duration (hours), often through histograms, probability density functions, scatter plots, or other visualizations. This approach, discussed by ESIG (2021), EPRI (2022), and Stenclik et al. (2022), and illustrated in Figure 7, provides a clearer picture of individual shortfall events across the entire distribution of event duration.

According to EPRI (2022):

> Many existing resource adequacy metrics can be adjusted to provide a more complete analysis of system risk. For example, evaluating the distribution of existing resource adequacy metrics, rather than only the average result (so called "full distribution metrics"), can provide additional insights. This is particularly helpful in the evaluation of outlier tail-

FIGURE 7

Reporting a Distribution of the Duration of Shortfall Events, Case Study of the Public Service Company of New Mexico



The figure shows a histogram of loss-of-load event duration (x-axis) across two resource adequacy scenarios that vary resources and load due to different levels of electrification (colored bars). Results show that even with two cases showing the same loss-of-load expectation, event duration can become longer in a high electrification scenario.

Source: GridLab (2023).

end events, which have a disproportionately large impact on the power system. Similarly, calculating existing metrics over a variety of time horizons (i.e., sub-annual metrics) can provide additional insights on the timing of adequacy events.

Reporting a distribution rather than a single expected value can convey a broader interpretation of resource adequacy risk. But while presenting these distributions is informative for planners, it does not directly translate into a minimum resource adequacy criterion for planning new resource procurements. In other words, these full distribution visualizations, while insightful, cannot easily be converted into a specific capacity requirement.

Reporting a distribution rather than a single expected value can convey a broader interpretation of resource adequacy risk, although it does not directly translate into a minimum resource adequacy criterion for planning new resource procurements.

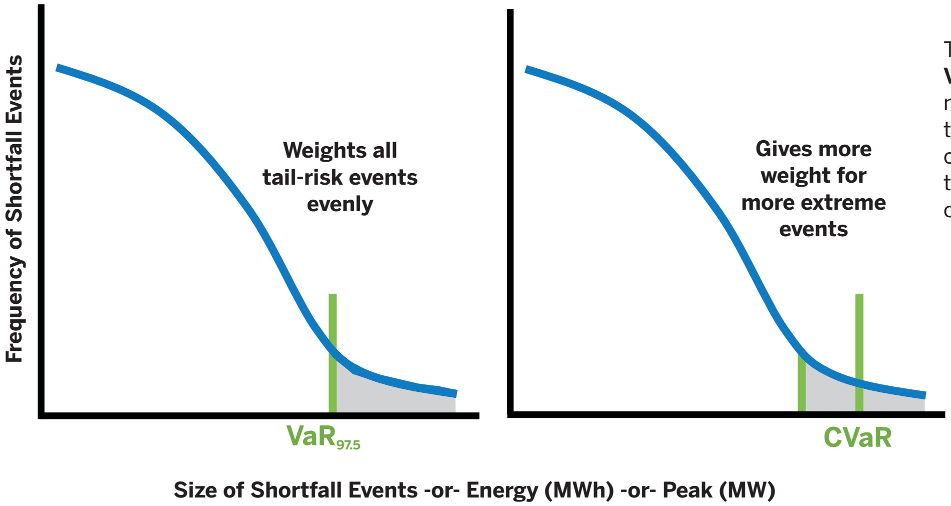
Directly Embedding Tail Risk Metrics into Planning Criteria

To directly embed tail risk metrics into planning criteria, specific metrics like value at risk (VaR) or conditional value at risk (CVaR) are sometimes used to calculate metrics that look at the worst-case events. A duration VaR-N, for example, collects the duration of the longest shortfall event for each simulation year, and then calculates the Nth percentile of the distribution of these durations from all the simulation years. A threshold imposed on the duration VaR-N indicates a criterion that any shortfall event with a duration longer than the threshold should not happen more often than once in y years, where y = 100/(100-N). A threshold on the "duration VaR99," for instance, indicates that duration exceeding that threshold should occur no more than once in 100 years. A threshold on this metric, such as limiting the duration VaR99 to less than 10 hours, means that a shortfall longer than 10 hours should occur only

FIGURE 8

Illustration of the Value at Risk (VaR) and Conditional Value at Risk (CVaR) Metrics

The **Value at Risk (VaR)** metric notes the size or duration of the largest and longest shortfall event for each simulation year (including zero for years with no events) and notes the Nth percentile of observations.



The **Conditional Value at Risk (CVaR)** metric calculates the average of the observations above the N^{th} percentile of observations.

The figure illustrates how the value at risk (VaR) (left) and conditional value at risk (CVaR) (right) metrics are calculated based on a distribution of resource adequacy samples. In this example, VaR is calculated as the value at the 97.5th percentile of all samples, whereas CVaR takes the *average* of samples above the 97.5th percentile. In both cases, only the tail end of the samples is considered. This can be measured across size (MW, MWh) or duration (hrs), as shown on the x-axis.

Source: Energy Systems Integration Group.

once in every 100 years across all of the simulations studied. Organizations like NWPCC are developing new criteria incorporating VaR metrics to limit tail risks related to shortfall event duration (hours), peak-hour loss (MW), and energy loss (MWh) (NWPCC, 2024).

CVaR, a variation of VaR, calculates the expected value of observations above the N^{th} percentile. It captures both the probability of a large event and the severity, by taking a weighted average of tail events into account rather than just reporting the N^{th} percentile value, as the VaR-N metric does. A duration CVaR-N, for example, starts with collecting the duration of the longest shortfall event for each simulation year, which results in a distribution of this type of duration from all the simulation years. Then the CVaR-N calculates the average of the subset of durations that are at or higher than the N^{th} percentile of the distribution. The comparison of these two metrics is provided in Figure 8.

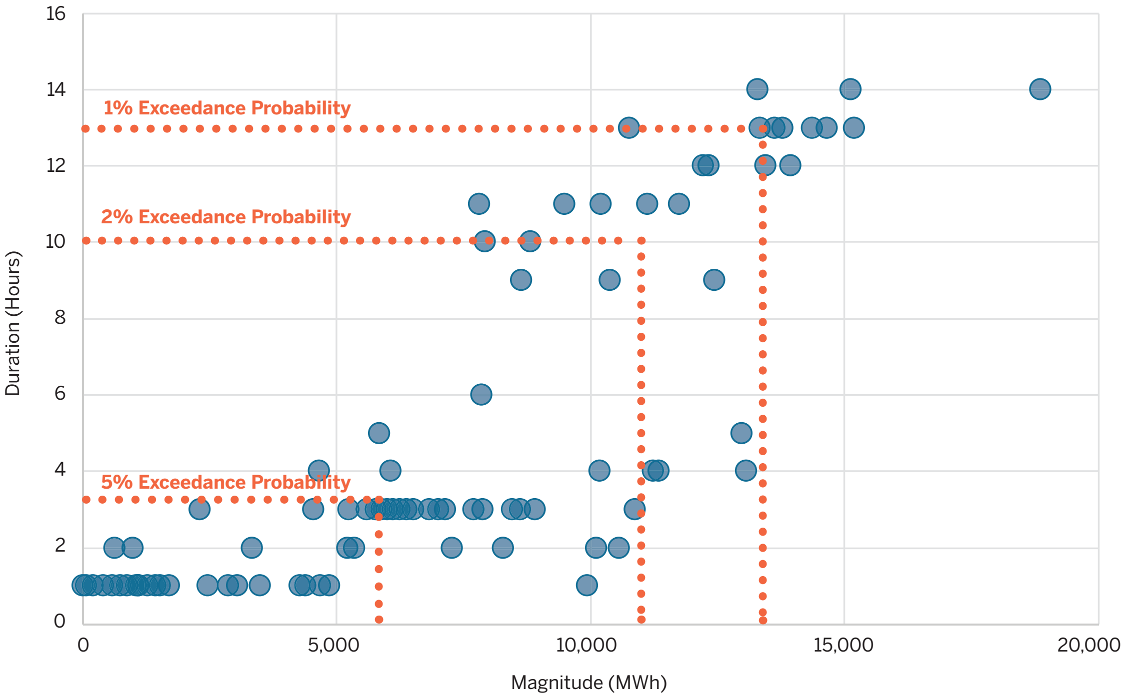
ERCOT is proposing a similar method, referred to as exceedance probability, to calculate these metrics (ERCOT, 2023). The process involves first ranking all events by magnitude and duration, then selecting an exceedance probability and determining the ranking that corresponds to this probability to set the risk tolerance thresholds. A visualization of this exceedance probability (or VaR) is provided in Figure 9 (p. 28), which shows how different exceedance thresholds can be drawn across multiple metrics.

Limitations of Tail Risk Criteria

However, there are limitations to consider in these tail risk metrics. For systems with substantial amounts of storage or other energy-limited resources, the duration and maximum size of events can be heavily influenced by how these resources are scheduled and optimized in resource adequacy simulations. Moreover, the use of metrics like VaR or CVaR requires confidence in the representation of probability distributions of future variables. Because resource adequacy events are hard to assign probabilities to (because they are so rare), VaR and CVaR metrics are equally imprecise. The stability and accuracy of these distributions, and their ability to capture all relevant future uncertainties, including those driven by climate change, are crucial for the effectiveness of these metrics.

FIGURE 9

Exceedance Probability Method for Magnitude and Duration



The figure illustrates a scatter plot of individual loss-of-load events from a probabilistic resource adequacy simulation of the ERCOT system. The events are plotted by the total unserved energy in each event (x-axis) and the event duration (y-axis). The dashed lines represent the exceedance levels, so 1% of all events are greater than 13,500 MWh in size and 13 hours in duration.

Source: Electric Reliability Council of Texas (2023).

Stress Testing for Extreme Events

Combining Stress Testing with Probabilistic Criteria

In some cases, the limited availability of data to confidently determine the probability of extreme events necessitates discrete analysis, or stress testing, rather than relying solely on statistical measures. The inclusion of probabilistic metrics in the planning criteria (like VaR or CVaR), while important, may not be sufficient to ensure system adequacy against rare, high-impact, low-probability events. This stems partly from the challenge in assigning probabilities to tail risks. Using the analogy of a "100-year flood," it's apparent that defining the probability and severity of a 1-in-100-year event is challenging, especially under the influence of climate change. The same is true for resource adequacy shortfalls on the power system, especially as system capability is becoming increasingly dependent on weather, which is affected by climate change.

Limited data are available to determine with confidence the probability of extreme events. This may require discrete analysis or stress testing rather than a statistical measure.

In addition, probabilistic resource analyses and resource adequacy criteria are often designed to determine how much capacity we need and how much investment is required locally. However, because the resource adequacy of the power system is often stressed by extreme weather, adding new capacity in the same region (be it gas, wind, solar, etc.) will likely be affected by the same system stressors. Simply put, adding more resources to the same region affected by the extreme weather may not be the most efficient way to improve resource adequacy. Therefore, a broader set of mitigations—beyond just capacity additions—may be warranted.

Deterministic stress-testing evaluates the power system's resilience in specific scenarios, such as a wide-area heat wave, a winter cold snap with limited gas supplies, or a multi-day renewable drought. These events can be explicitly modeled, allowing planners and regulators to understand system risks and prioritize mitigations beyond simply adding new capacity, and offering insights into system vulnerabilities that probabilistic resource adequacy assessments might overlook. An illustration of this dual approach of combining probabilistic analysis alongside deterministic stress-testing is shown in Figure 10.

In this report, stress testing is intentionally included in the resource adequacy framework. The reason for this is three-fold. First, it emphasizes the importance of considering extreme events. Second, it ensures that stakeholders consider a wide range of resource mitigations, not just additions to capacity. The final reason is that if these events continue to fall outside of the resource adequacy framework, there is limited ability to invest or procure resources to ensure resource adequacy.

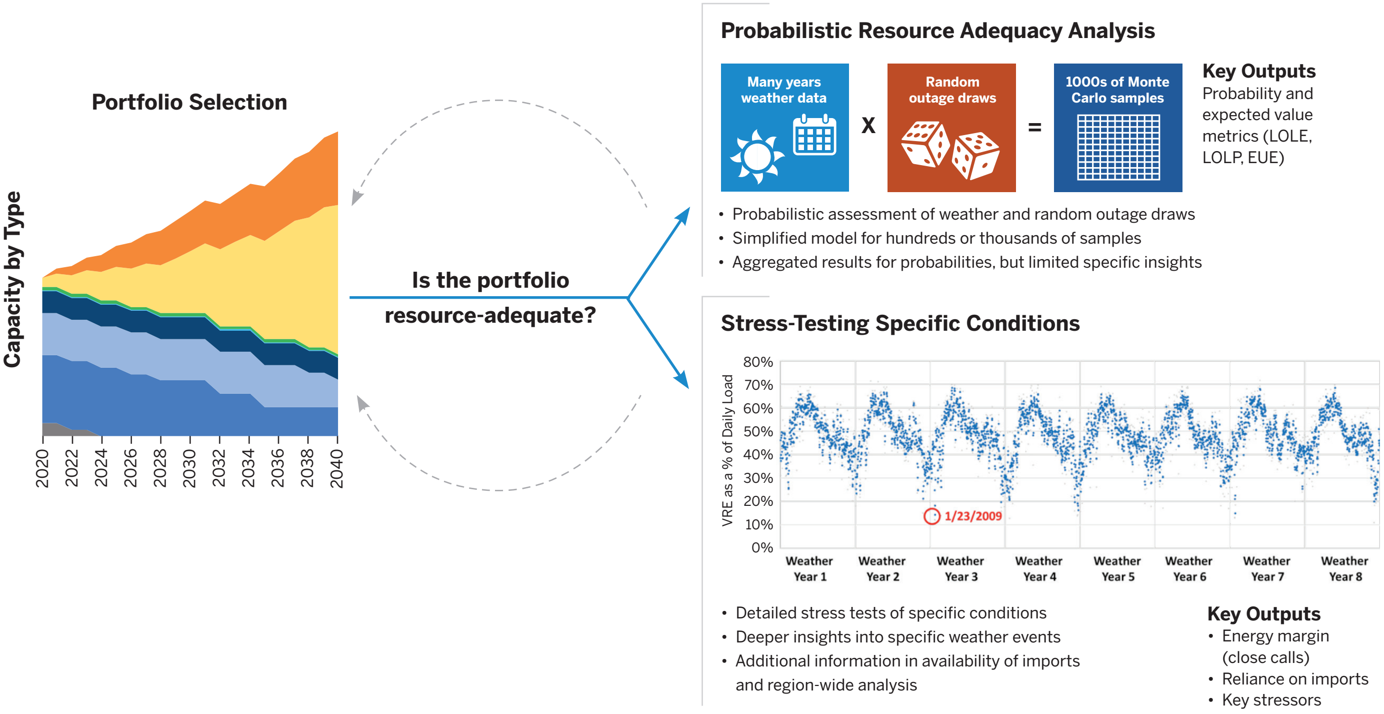
There is a blurry line differentiating resource adequacy from resilience. It is a topic of debate and an evolving discussion among planners and regulators. On the one hand, some planners believe that specific natural disasters —among which extreme weather is included—fall outside of the resource adequacy umbrella. Others argue that avoiding loss of load is the objective, regardless of the underlying cause. By expanding the resource adequacy criteria to include stress testing, planners can incorporate a wider range of risks and mitigations to ensure reliability.

Today, some planners are starting to incorporate stress testing in their portfolio analysis. However, where this type of deterministic stress-testing analysis is conducted today, it is often done without established methods and is done in a "one-off" analysis with limited interactions with other planning or resource procurement efforts.

There are two primary approaches to identifying periods to stress test:

FIGURE 10

Combining Probabilistic Resource Adequacy Analysis with Stress Testing



The figure illustrates the two-pronged approach of determining resource adequacy of a resource portfolio, one that includes probabilistic resource adequacy analysis and one that selects challenging time periods for a deterministic stress-testing approach.

Source: GridLab (2022).

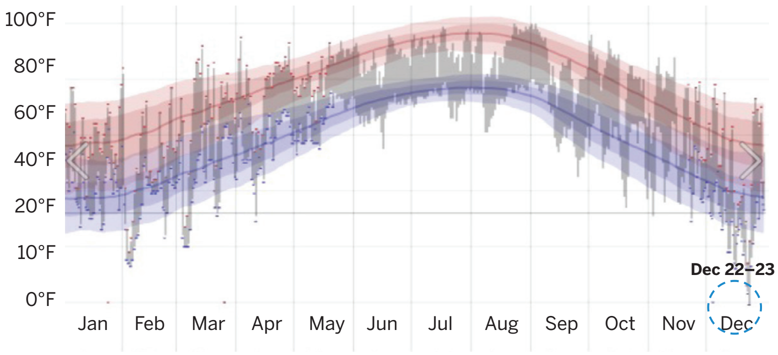
FIGURE 11

Daily Loss-of-Load Probability Metrics to Screen for Periods of System Stress

ERCOT Top 5 LOLP Days—High Renewable Portfolio

Weather Year	Date	Daily Min/Max Temp (Austin)	Daily LOLP (%)
1982	Jan 11	11 / 30°F	76%
1989	Dec 22	12 / 23°F	100%
1989	Dec 23	4 / 29°F	72%
1985	Feb 2	14 / 30°F	100%
1985	Feb 1	19 / 23°F	96%

1989 Austin Temperature



The figure shows the top five days of loss-of-load risk across a 40-year historical period of the ERCOT system. In a higher-renewable future, loss-of-load risk was concentrated in a select few days with anomalous weather, and loss-of-load probability reached 100% in some cases, representing a potential stress-test condition to evaluate further.

Notes: ERCOT = Electric Reliability Council of Texas; LOLP = loss-of-load probability.

Sources: EPRI (2023) and WeatherSpark.com.

- **Inspection of probabilistic analysis results:** Periods of higher-than-normal risk can be identified by examining daily loss-of-load probabilities from probabilistic resource adequacy analysis. Assigning each day a loss-of-load probability (LOLP) can identify periods when there is elevated risk in the historical weather record.

  For example, a recent assessment of resource adequacy in ERCOT showed that in the current resource mix, resource adequacy risk is largely confined to the summer, mid-day, peak (gross) demand period with 30% of all loss-of-load events occurring during the June 25-26, 2012, weather year's heat event (EPRI, 2023b). In the future, however, the assessment showed that risk shifts to the winter, and nearly half of all loss-of-load events occur in just three days of the 40-year historical weather record (Figure 11). Furthermore, these weather events approached 100% LOLP, indicating that the system would have a resource adequacy shortfall—requiring emergency actions or resulting in involuntary load-shedding—if the weather event were to occur. Since these events occur outside of a commonly used 1-day-in-10-years probability, they warrant stress testing of specific risks even if their probabilities cannot be accurately quantified.

- **Stakeholder input for identifying critical risks:** A second approach is to identify risk periods that may not be included in the probabilistic dataset. Stakeholders—including planners, regulators, and others—can help identify critical periods and types of risk, even if they do not appear in the probabilistic analysis. This approach ensures that the resource mix is robust against key adequacy events even if they are not captured by the probabilistic analysis. An example would be assessing the impact of a repeat of a historical weather event combined with other system stressors, providing a realistic, understandable, and deterministic "event."

The credibility of weather scenarios under consideration is crucial. Planners should avoid creating implausible doomsday scenarios, which may be overly conservative (leading to increased costs for mitigations), and instead focus on credible, albeit rare, weather events or other stressors.

Stress testing in this context allows for a more detailed approach and a wider range of sensitivities, offering better insights into key risk drivers than a probabilistic approach used today, which considers thousands of statistical data

points, making it difficult to interpret or understand the root causes of adequacy events. Stress testing can complement probabilistic methods in the planning criteria by providing detailed insights into system resilience against specific high-impact, low-probability events. This approach helps planners and operators understand the nuances of risk in practical, relatable terms, guiding decisions to enhance reliability.

Including Stress Testing in the Criteria and Evaluating a Wider Range of Mitigations

Inclusion of stress testing in resource adequacy criteria involves combining both probabilistic methods across a wide range (thousands) of different system stressors along with a deterministic evaluation of system events that are most concerning to planners. For instance, multi-part planning criteria could include, in addition to a probabilistic assessment that measures LOLE, LOLH, and EUE, a secondary assessment of a deterministic stress event. If this assessment fails (i.e., unserved energy is seen to exceed a certain magnitude in the planning studies), it could trigger additional investments or deferred retirements to maintain reliability, even if the probabilistic 1-day-in-10-years threshold is met.

Examples of this method are currently employed by two load-serving entities in Colorado, Tri-State Generation and Transmission Association (Tri-State) and Public Service of Colorado (PSCo), in their long-term planning processes (Tri-State Generation and Transmission Association, 2023). Tri-State's framework involves Level 1 and Level 2 adequacy metrics. The first level is a multi-metric criterion that adheres to the standard 1-day-in-10-years LOLE criterion, in addition to planning reserve margin and EUE targets. The second level assesses whether the system can withstand specific extreme weather events. In this second level, compliance is measured using additional metrics, such as limiting loss-of-load hours to no more than three in any extreme weather event and ensuring that unserved energy does not exceed 20% of load in any hour.

In restructured markets, the stress-testing approach could be used to increase capacity requirements, establish other grid service products, or set up a reliability backstop procurement. Some of this is already practiced ad hoc to ensure reliability during extreme events—like requiring dual-fuel capability for generators in some regions—but stress testing is often not directly included in the resource adequacy criteria, which drives procurement decisions using today's probabilistic LOLE metric.

A critical benefit of pairing a probabilistic metric with stress testing is that options available to improve reliability in a stress-test event might not be identified solely by evaluating additional new resources, as is typically done when using probabilistic resource adequacy analysis. Since stress testing is based on a deterministic approach —only considering a single set of challenging grid conditions—it should consider non-capacity options, including external assistance from neighboring regions, modifications to plants to allow operation during extreme weather conditions, transmission reinforcement, or specific demand-response programs.

Multi-metric Criteria

Objectives of Multi-metric Criteria

Given the evolving dynamics in resource adequacy analysis, the changing energy resource mix, and consumer preferences for reliability, adopting a multi-metric criteria approach may be prudent because it provides a more comprehensive assessment of the size, frequency, and duration of shortfalls; explicitly considers tail risks; and can stress-test extreme events that may fall outside historical records. A multi-metric approach can support the following five objectives.

Limit the Likelihood of Shortfall Events

First and foremost, multi-metric criteria should be designed to continue limiting the likelihood of shortfall events. This aspect is analogous to the current framework, including the 1-day-in-10-years LOLE metrics. Maintaining this component is vital, as it offers a baseline understanding of how often the power system might experience unserved load. A high frequency of shortfall events (i.e., greater than the 1-day-in-10-years threshold) implies that the system is uneconomical and would rely too often on expensive emergency measures to keep the lights on.

No one metric is the solution; multi-metric criteria are needed to consider size, frequency, timing, and duration of shortfalls.

Address Low-Probability, High-Impact Events

A second critical objective of the multi-metric approach is to support mitigating the risk of low-probability, high-impact events that could have disproportionate effects on human life and economic damages. These tail risks, often rare but catastrophic, necessitate special attention to prevent extensive disruptions and losses. This can be done through both statistical measures (i.e., VaR, CVaR) and through deterministic measures like stress testing.

Embrace a Flexible, Multi-dimensional Approach

In addition, the multi-metric criteria help ensure that regulators and planners adopt a flexible approach to resource adequacy that acknowledges multiple dimensions of risk. This approach involves looking beyond capacity procurements to a broader set of potential mitigations. It encompasses various strategies, including infrastructure resilience enhancements, demand response programs, and diversification of energy sources, to address different types of resource adequacy risks effectively. A multi-metric approach can help elucidate the unique contributions of these mitigations toward resource adequacy, for example, by giving planners visibility into whether a given solution may be more effective in reducing the depth or duration of outages during rare tail events, or is better suited to reducing the likelihood of outages during less severe but more frequent at-risk hours.

Inform Stakeholders of Associated Risks

Finally, an essential function of a multi-metric approach is to provide all parties, from system operators to regulators to emergency planners (officials responsible for responding to emergency events) with comprehensive, but easy to understand, information about the risk to the power system. This information is crucial not only for power system planning and investment but also for preparing emergency services and planning behind-the-meter back-up generation needs for critical infrastructure.

To revisit the flood analogy, for a community planner it's not enough to know the frequency of flooding. It is equally important to understand the implications for the community, including the event's duration and potential damages. Similarly, in power system planning, these additional metrics and event descriptors are crucial for devising appropriate responses to resource adequacy events.

Identify and Limit the Most Impactful Risks for a Given System

Risk profiles and the impacts of different types of outage events vary greatly from region to region. As a first step in thinking about adequacy with a changing resource mix, planners and policymakers can identify the conditions and extreme events that pose the greatest risk and most significant consequences for their region. This could be outages during summer heat waves in the Southwest or widespread winter outages resulting from extreme cold events and gas supply disruption in the Midwest. The multi-metric criteria should be tailored to accurately capture the risks of greatest concern.

Options for Multi-metric Criteria

There are different options for implementing multi-metric criteria in power system planning and procurement decisions. Entities in charge of power system adequacy must decide what each criterion should be used for. Examples include: (1) determining whether the system is adequate, (2) establishing an effective capacity need and a planning reserve margin, and (3) accrediting resources to measure the individual resource adequacy contributions they provide. It's also possible that a region may use a single metric for near-term markets and procurement (i.e., next year's capacity auction or integrated resource plan procurement) but use multiple metrics for longer-term planning. This approach can take different forms, two of which are outlined below.

Option 1: Use a Primary Metric for Accreditation and Procurement, with Supporting Metrics

In this option, entities responsible for resource adequacy continue to use a single metric for establishing the capacity requirements and for individual unit accreditation, but additional metrics, and data visualizations (i.e., histograms and scatter plots), are used to assess overall system risk. This is the least disruptive change to the current, single-metric, LOLE resource adequacy criterion and is already in use in many jurisdictions. At a bare minimum, responsible entities should show the distribution of the metrics (Figure 7, p. 25) and communicate statistical measures like VaR and CVaR. It is crucial for planners to properly explain this approach to regulators and stakeholders and to ensure that risk is properly mitigated, rather than just looking at a single, expected value metric (the primary metric) and disregarding the supporting metrics.

Option 2: Implement Multi-metric Criteria, and Require All Metrics to Meet the Criteria

This establishes a true multi-metric approach and tests the portfolio across each metric. In this example, while all metrics must be satisfied (i.e., be below the adequacy threshold) for a system to be considered adequate, a single metric in the criteria becomes the binding constraint in a given simulation or accreditation process. This "least common denominator" becomes the effective planning criterion until the system changes and another metric becomes binding. For example, a four-metric criteria may result with one metric violated, such as frequency or magnitude, but the rest satisfied. However, as the resource mix changes, one of the additional metrics could later become the binding constraint in the modeling, allowing the multi-metric criteria to automatically adjust as the power system risk changes. This ensures that the system is robust against different facets of risk.

An example of this is currently being proposed by NWPCC and is illustrated in Table 6. In this example, the system needs to meet a four-part criteria, combining both expected and tail-end approaches. The metrics include a maximum frequency using loss-of-load events (LOLEv), a maximum tail-end threshold for duration (hours), a maximum tail-end threshold for one-hour peak (MW), and annual energy (MWh) shortfalls. Thus, this approach explicitly takes into account the size, frequency, and duration of shortfalls. The NWPCC is undertaking an extensive stakeholder engagement process to develop the metrics and thresholds.

Dynamic Resource Adequacy Requirements

The evolving nature of the power system necessitates the establishment of a dynamic resource adequacy framework—one that is adaptable to changes in resource mix and load and revisited over time, either from year to year or on a regular planning cycle. Relying solely on a static metric (LOLE or LOLH) is shortsighted given the

TABLE 6

Summary of Proposed New Adequacy Metrics, Northwest Power and Conservation Council

Metric	Units	Objective	Definition	Interim Threshold
LOLEv winter & summer	Events/year	Prevent overly frequent use of emergency measures	Loss-of-load events = expected number of shortfall events per year (total number of shortfall events divided by the total number of simulations)	0.1 in summer 0.1 in winter
Duration $VaR_{97.5}$	Hours	Limit the risk of long shortfall events to 1/40 years	97.5th percentile of the distribution of longest shortfall events for all simulations	8 hours
Peak $VaR_{97.5}$	MW	Limit the risk of large capacity shortfall to 1/40 years	97.5th percentile of the distribution of highest single-hour shortfall for all simulations. Also reporting Normalized Peak VaR 97.5.	1,200 MW (corresponding to 3% of maximum load)
Energy $VaR_{97.5}$	MWh	Limit the risk of large energy shortfalls to 1/40 years	97.5th percentile of the distribution of total annual shortfall energy for all simulations. Also reporting Normalized Energy VaR 97.5	9,600 MWh (corresponding to 0.0052% of annual load)

The table shows the proposed thresholds for the four-metric criteria developed by NWPCC. While the table includes specific thresholds ("interim threshold"), these should be used as examples, not necessarily the appropriate threshold for other regions, as these may be region-specific.

Source: Northwest Power and Conservation Council (2024).

rapid transitions underway in the energy sector. The evolution toward a need for more dynamic resource adequacy criteria is driven by increased generator retirements, the adoption of variable renewables, increased storage deployment, and enhanced load flexibility.

Additionally, as electrification extends to industrial processes, transportation, and heating, more customers will be reliant on the power system, and that reliance intensifies as demand for electricity increases greatly. As new end uses electrify, customer preference may require the resource adequacy criteria to adjust as well. However, these new electrified loads also present opportunities for load flexibility, such as vehicle-to-home back-up, distributed batteries, and heat pumps, which tends to shift some of the responsibility for reliability away from the bulk system to individual customers or groups of customers. Targeted load-shedding could further mitigate the need for heightened reliability, emphasizing the economic aspect of reliability decisions.

Embracing multiple metrics is crucial for covering this evolving risk landscape, but it is equally important to allow criteria and thresholds to adjust with a changing resource mix and consumer preferences. As we navigate this transition, it is imperative to treat the establishment of new criteria as an ongoing, adaptable process, fully embracing the inherent uncertainties of a changing energy landscape.

Embracing multiple metrics is crucial for covering the evolving risk landscape, but it is equally important to allow criteria and thresholds to adjust with a changing resource mix and consumer preferences.

Additional Considerations for Multi-metric Criteria

Developing multi-metric criteria for resource adequacy presents several challenges and considerations that must be addressed to ensure its effectiveness and practicality. While a multi-metric approach offers a more comprehensive understanding of system adequacy, its implementation needs to be carefully balanced to avoid undue complexity or burden for planners, regulators, and stakeholders.

Need for Transparency and Communication with Stakeholders

The first issue is related to transparency and communication, and the risk of making an already complex process more opaque and creating barriers to procedural equity. The existing single-metric, LOLE or LOLH criterion, prevalent in most regions, is already difficult to interpret. For many stakeholders, especially those not regularly engaged in power system modeling and probabilistic analysis, understanding and interpreting these metrics can be daunting. Introducing multi-metric criteria could further exacerbate this issue, making it more difficult for stakeholders to grasp and engage effectively in the planning process. To overcome this, engagement and education with a wide range of stakeholders would need to occur early in the process when reconsidering the resource adequacy criteria.

Risk of Defaulting to a Single Criterion

Another potential challenge with multi-metric criteria is the risk of it defaulting to a single metric if one consistently becomes the binding one within the set. If all metrics are given equal weight, but one metric consistently becomes binding first, it could inadvertently become the new de facto resource adequacy criterion. This situation could undermine the intent of having a multi-dimensional approach to resource adequacy. To address this challenge, planners can clearly articulate which criterion is binding—and the time period it is binding in—as these can, and should, change throughout the planning horizon.

New Considerations for Procurement

The traditional single criterion was easily translated to planning reserve margins and capacity procurement decisions. With multi-metric criteria, these equivalencies are less clear. Some resources may contribute to mitigate the frequency of outages but not their duration, while other resources may provide significantly higher capacity in certain seasons compared to others, with their ability to mitigate outages depending on the timing. The continued use of the planning reserve margin to reflect capacity needs will be increasingly challenged with a diverse set of criteria and resources. Similarly, procurement decisions will have to be more tightly coupled with the resource adequacy assessment to ensure they are responsive to the different metrics. If the planning criteria is changed, it should be consistent across various stages of the procurement process, including setting the capacity threshold and accrediting individual resources.

Need for Careful Consideration of Storage Dispatch

A fourth challenge is that the power system is becoming increasingly energy-limited, with higher levels of storage and load flexibility. The scheduling of these resources can significantly influence which metric becomes binding. This necessitates careful consideration of storage dispatch in the implementation of multi-metric criteria. Planners and regulators need to ensure that the scheduling and optimization of these resources in the resource adequacy modeling is done as it would be in practice, without optimizing it to meet a specific adequacy criterion. In markets where various types of storage can be procured using different types of contracting mechanisms, each of these can have implications that affect resource scheduling. This can add challenges both for planners to model the best optimization strategies and for regulators to decide what objective they prioritize the most. Planners should be clear and intentional in how storage is scheduled in resource adequacy studies.

In summary, while the use of multi-metric criteria for resource adequacy represents a step forward in capturing the complexities of modern power systems, its implementation comes with challenges that need careful consideration. The set of successful adoption of this approach requires balancing complexity with procedural equity and access to information, ensuring that all metrics meaningfully contribute to decision-making, and adapting the framework to suit energy-limited systems.

The successful adoption of a multi-metric criteria approach requires balancing complexity with procedural equity and access to information, ensuring that all metrics meaningfully contribute to decision-making, and adapting the framework to suit energy-limited systems.

Incorporating Economics in the Resource Adequacy Criteria

Adequacy Is an Economic Trade-off

Establishing the appropriate level of power system adequacy is not a straightforward task; it requires a collective judgment call involving planners, regulators, customers, and other stakeholders. Resource adequacy in power systems should not be viewed as black and white or a line in the sand. Contrary to the prevalent notion in current planning practice, power system adequacy is not a binary indicator where the system is deemed either adequate or not. Instead, resource adequacy represents a continuum of risk. Since no utility in the world plans for a 100% reliable system (due to exorbitant costs), the risk of shortfall is never zero. Instead, resource adequacy should be viewed as a continuum in which the level of risk is dependent on the amount of resource investment. The appropriate level of adequacy is determined by the trade-off between what customers are willing to pay versus how much shortfall they are willing to tolerate.

The objective of the resource adequacy criterion is to strike the right balance between reliability and cost. Setting the criterion too high can lead to prohibitively high investment costs, while setting it too low risks diminished reliability and the potential for significant economic damages. It's crucial that this intrinsic link between cost and reliability is transparent and well understood by all involved parties.

The resource adequacy criteria should be used to establish the appropriate trade-off between reliability and cost. Cost and reliability are intrinsically linked, and this trade-off should be clear.

Because assessing the cost of damages is challenging (for instance, knowing the value of lost load, or VoLL), many system planners and regulators set resource adequacy thresholds without adequately weighing the associated costs. This approach can lead to inefficiencies and suboptimal investment decisions. There are instances when enhancing reliability may come at little or no additional cost, making the investment clearly worthwhile. Conversely, there are situations when a small increase in reliability comes with a substantial price tag. In such cases, it may be prudent to accept a higher level of risk, especially if the potential damage from the incremental loss of service is manageable. The challenge in setting resource adequacy thresholds lies in finding the appropriate balance between cost and risk.

Setting the Threshold

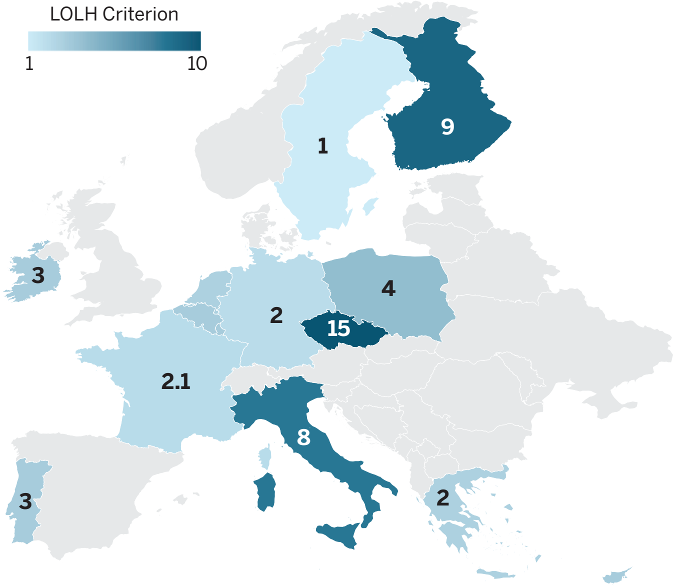
Thus far, this report has primarily explored which metrics should be included in the resource adequacy criteria and how the set of criteria is used to drive investments. But it is equally important to determine the level at which the resource adequacy criteria should be set. The level or threshold for an adequacy metric represents the maximum allowable value for that metric. For example, a commonly used threshold for LOLE is 0.1 day per year, meaning that the system is deemed adequate if LOLE is not greater than 0.1 day per year. The choice of threshold —whether it be planning for a 1-day-in-10-years, 1-day-in-5-years (indicating a less reliable system), or 1-day-in-20-years (indicating a more reliable system)—has profound implications for the investments required in the power system. The regulatory decisions setting thresholds significantly impact resource investment and retirement decisions within a power system. The challenge lies in establishing how stringent these resource adequacy requirements should be, because they impact not just the cost of electricity but, by extension, the economic and social wellbeing of consumers. This section provides a general discussion of how adequacy thresholds can be determined while considering costs and economic decisions.

Regardless of the resource adequacy criteria employed, three critical questions are: (1) how to establish these thresholds, (2) who should be responsible for making these decisions, and (3) how to balance cost and reliability objectives. In many parts of North America, the prevalent 1-day-in-10-years LOLE threshold is often adopted without clear justification, with regions using it mainly because it is widely accepted.

In Europe, for example, while there is a single LOLH metric used for the criterion, each country sets its own threshold using a common method and is free to recalculate that threshold, if necessary, over time (Figure 12). This approach demonstrates that a common framework can be used without necessarily requiring uniformity in threshold levels, and these thresholds do not have to be static over time. In the European example, LOLH thresholds are set using VoLL and the cost of new entrant (CONE), which are expected to change over time, consequently leading to new thresholds as the system changes.

FIGURE 12

Select European National Reliability Standards as Applied by EU Member States as of June 2023



The map shows the various LOLH across select European member states. While the same methodology is used to develop the criterion in each member state, the threshold can vary due to local value of lost load (VoLL) and cost of new entry (CONE) and range between 1 and 15 hours per year. Note that the European criterion is expressed in LOLE (hours/year), which equates to LOLH used in North America and used in this report.

Notes: LOLE = loss-of-load expectation; LOLH = loss-of-load hours.

Source: Energy Systems Integration Group; data from the European Union Agency for the Cooperation of Energy Regulators (2023).

Options Available for Incorporating Economic Principles

There are at least four options to incorporate economic principles into the resource adequacy criteria:

- Increase transparency
- Ensure proper price formation during scarcity conditions
- Calculate the investment vs. damages ratio
- Identify the economic optimum

Increase Transparency

At a minimum, transparency in the economic-reliability trade-off should be provided. In vertically integrated utilities with integrated resource planning, future resource portfolios can be developed, and costs calculated, using different levels of the criterion (e.g., 1 day in 5 years, 1 day in 20 years). This approach would clearly show the reliability-cost trade-off, and regulators and stakeholders could interpolate both cost and resource mix changes across different levels of adequacy. Using this information, the marginal cost of resource adequacy (measured in $/day of shortfall per year) could be established, and the marginal resource adequacy resource(s) to increase power system adequacy can be identified.

In restructured markets with capacity auctions, the auction results can be provided at various levels of capacity requirements, providing a “shadow price” for increased resource adequacy—or by using a downward-sloping demand curve. This would show what the market clearing price would have been had the resource adequacy threshold been higher or lower, again providing stakeholders with information about both the marginal cost of adequacy and the marginal resource necessary to achieve different levels of reliability.

This option is strictly for cost-effective planning. It does not provide a way to determine the appropriate minimum threshold; it simply reflects the resource cost associated with providing varying levels of reliability for different resource mixes. This provides transparency to the "cost of investment" versus the "level of risk" relationship but does not necessarily yield a resource adequacy target.

Ensure Proper Price Formation During Scarcity Conditions

Another method to clearly link the economic and reliability trade-off is through a real-time scarcity pricing (i.e., operating reserve demand curve) framework. In this example, prices increase substantially when capacity reserves on the system become tight, affording a clear price signal to the market. These price signals allow customers (loads) and suppliers (resources) to align with willingness to pay for reliability, increase load participation, and constitute a price signal for new capacity to enter the market and improve resource adequacy if it is efficient to do so. This also creates an incentive for new demand response and transparent hedging to reduce risk (and thus price volatility). This approach is not a minimum criterion, nor does it provide a long-term price signal, but clearly links power system reliability and economics using a price signal.

Calculate the Investment vs. Damages Ratio

In the European Union, the methodology for the European Resource Adequacy Assessment as developed by the European Network of Transmission System Operators for Electricity (ENTSO-E) and approved by the European Union Agency for the Cooperation of Energy Regulators (ACER) establishes the reliability standard by combining a VoLL with CONE (ACER, 2023). The LOLH threshold can then be calculated by dividing the CONE (\$/MW/year) by the assumed VoLL (\$/MWh). The resulting metric is the expected number of shortfall hours per year, which sets a reliability standard that is directly linked to two economic measures of investment and risk.

$$\text{LOLH (hrs/year)} = \frac{\text{CONE (\$/MW/yr)}}{\text{VoLL (\$/MWh)}}$$

An example output of this analysis is provided in Table 7 (p. 41), which was developed by the UK Department of Energy and Climate (although not a European Union member state, the UK uses this method as well). The table shows the implied LOLH threshold after taking into account the VoLL (rows) and the CONE (columns).

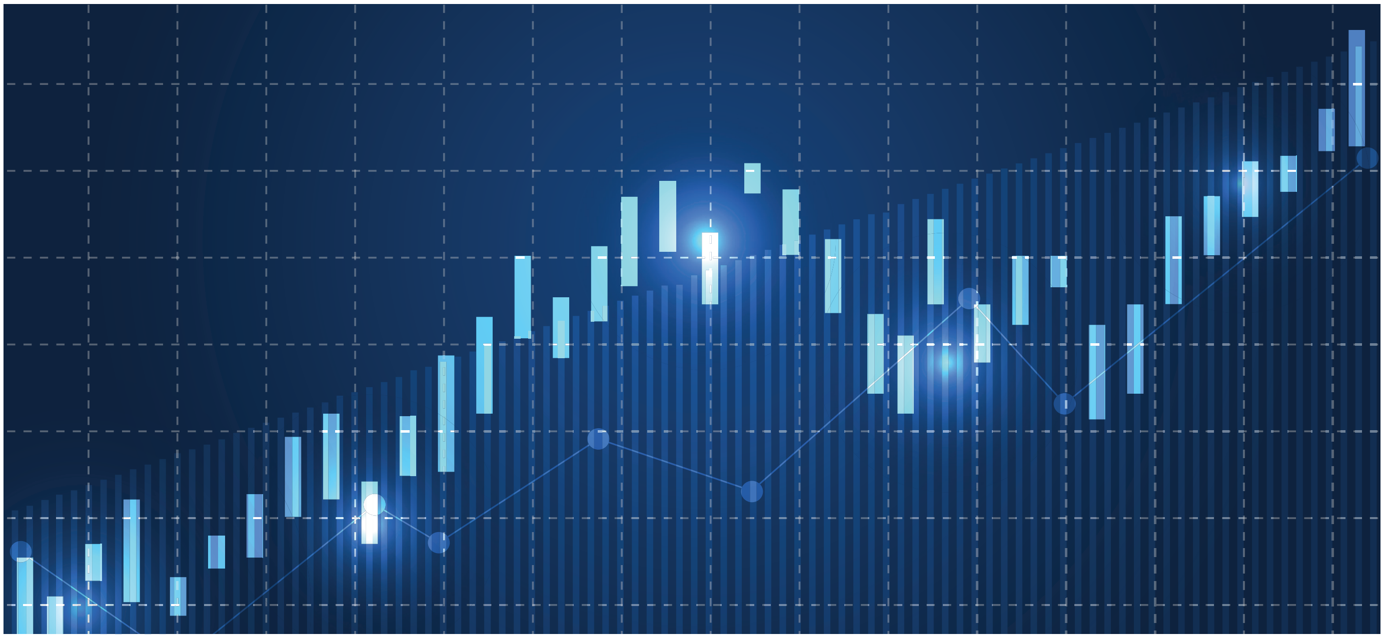


TABLE 7

Equilibrium Reliability Standard in LOLH, United Kingdom

VoLL (£/MWh)	Long-term CONE (£/kW)		
	Low	Central	High
35,500	0.90 hr	1.33 hrs	1.87 hrs
17,000	1.88 hrs	2.78 hrs	3.91 hrs
10,300	3.10 hrs	4.59 hrs	6.43 hrs

The table presents a range of LOLH criteria as a function of cost of new entry (columns) and value of lost load (rows), with a resulting range of 0.9 to 6.43 hours per year.

Source: UK Department of Energy and Climate Change (2013). [https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/]

The resulting LOLH thresholds range from 0.9 hours/year to 6.43 hours/year, depending on the economic assumptions made.

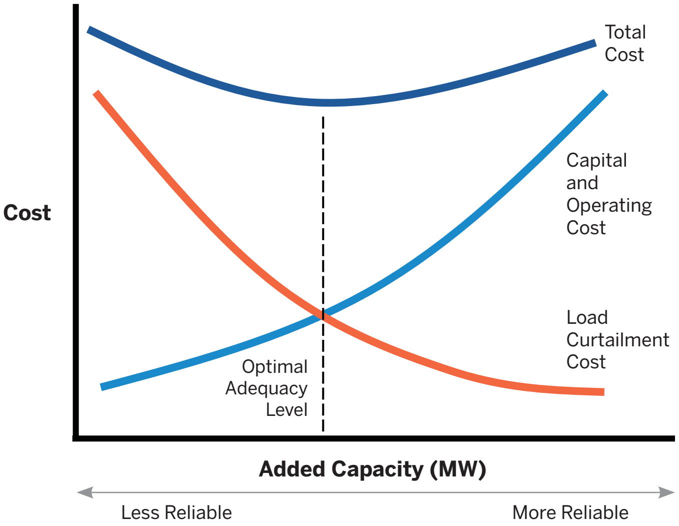
Identify the Economic Optimum

Finally, the most detailed, yet challenging, method to link economics and reliability in the resource adequacy criteria is to calculate the economically optimal level of adequacy. Similar to the investment-to-damages ratio, this method calculates the relationship between resource investment and damage (load curtailment) cost. Increasing resource investments will decrease load curtailment costs, thus improving resource adequacy. This inverse relationship is shown in Figure 13, which also shows total cost (investment cost plus curtailment cost). The optimal level of investment is the point along the horizontal axis (added capacity) at which total cost is the smallest. This point determines the optimal amount of resource investment to make. But perhaps more importantly, related to the resource adequacy criterion, the amount of expected unserved load at the optimal point becomes an EUE resource adequacy target, not a threshold. Planning to a lower or higher EUE value leads to more costly systems.

Using this method is challenging because it requires running many probabilistic analyses to create the investment cost (capital and operating cost) and damage cost (load curtailment cost) curves shown in Figure 13 (Billinton and Allan, 1996). In addition, each set of curves is generally based on a fixed VoLL and a fixed resource or resource mix. Thus, the optimal investment is dependent on the choice of investment resource(s), which means the corresponding target EUE is only valid for those resource(s). Thus, if this method were to be used to set the target EUE as the resource adequacy criterion, it would be prudent to choose a mix of new resource types that is most likely to be acquired.

FIGURE 13

Optimal Adequacy Level as a Function of Investment Cost and Load Curtailment Damages



The figure illustrates the trade-off between resource adequacy as a function of added capacity (x-axis) and cost (y-axis). As capacity is added to the system, cost (damages) from load curtailment decreases, but the capital and operating costs increase. The optimum level of reliability is where the sum of the two costs, representing total costs, is minimized.

Source: Energy Systems Integration Group.

There is a way to approximate the method described above. Since the optimal investment is likely to occur at or near the point where investment cost equals damage cost, a target EUE can be determined by equating the expression for the total net CONE (function of CONE) with the expression for societal cost (function of VoLL and EUE) and then solving for EUE.

In this example, several parameters must be assumed up front, including:

- **Value of lost load ($/MWh)**, a monetary indicator expressing the costs associated with an interruption of electricity

- **Societal cost (damages) of load curtailment ($/MWh)**, equal to VoLL times EUE
- **Cost of new entry (CONE, $/kW-year)**, the levelized capital and fixed costs for a new resource or portfolio of resources
- **Total net cost of new entry (net CONE, $/kW-year)**, the annual cost of adding a new resource—or portfolio of resources, accounting for all capital, fixed, and variable costs, and reductions in other system costs (e.g., displacement of higher-cost resources and revenue from out-of-region market sales)

Challenges with Value of Lost Load and Cost of New Entry

A significant challenge in incorporating economic principles into resource adequacy criteria is establishing assumptions for the VoLL and CONE. These two inputs are pivotal in balancing reliability with cost objectives, yet they present unique complexities in their practical assessment.

Establishing VoLL, which represents the societal economic damages from lost load, has long been a contentious and complex issue. Trying to establish a single VoLL is problematic, as different customers, with different preferences, can have a wide range in willingness to pay. It also raises important equity concerns, as socioeconomic factors could influence how an outage affects customers' wellbeing. This difficulty arises from the challenge in assessing the economic impact of power outages, which varies not only by customer class but also by season and time of day. Moreover, the impact is nonlinear, meaning that the cost of lost load escalates disproportionately with the duration and magnitude of the outage. Effectively, VoLL is a function of customer type, seasonality, duration, and outage severity, but is reduced to a single number in analysis. While challenging, it is essential to be clear about the perspective from which the cost is being measured—whether it's from the utility's standpoint or the customers'. Both cost functions are calculable and significant, similar to methodologies employed in the insurance and reinsurance sectors. The key is to avoid treating VoLL as a fixed value, as this would oversimplify the complex nature of the costs associated with power outages.

Relative to VoLL, assessing the CONE is somewhat more straightforward, yet it still presents challenges. CONE has traditionally been based on the cost of adding a natural gas combustion turbine, but this has been complicated by the diverse resource mix in modern power systems and public policies shifting away from fossil fuel infrastructure. The "new entrant" today will more likely be a portfolio of new entrants, including wind, solar, battery storage, and potentially gas resources, which together aim to achieve not just reliability but also economic and environmental objectives. This makes determining the marginal capacity resource or a combination of resources increasingly complex.

In summary, while integrating economic principles into resource adequacy criteria would be valuable for creating an economically efficient power system, it is fraught with challenges. Establishing accurate and representative values for VoLL and CONE requires careful consideration of various factors, including customer impacts, the diversity of the resource mix, and the evolving nature of power systems. Even then, the calculation of VoLL is inherently complex and contentious and CONE, while more straightforward, remains difficult. Regardless, addressing these challenges is necessary for the development of economically efficient resource adequacy criteria.

Implementation and Recommendations

Making the Switch: Options for Implementing Changes to Adequacy Criteria

Implementing changes to resource adequacy criteria within the power system is a complex process. It requires achieving a broad consensus among various stakeholders including utilities, grid operators, regulators, and customers. Such transitions should be approached with careful consideration, acknowledging the impact and scope of these changes.

This report, while not prescriptive, provides guidance to entities in charge of defining resource adequacy criteria. It lays out a need to move beyond a single-metric criterion to better capture size and duration of risk, to explicitly consider tail risks that have disproportionate damages, and to better link cost and reliability objectives.

A few select regions are pioneering a transition toward new resource adequacy criteria, and many others are considering taking action. In the United States, for example, the NWPCC proposes a four-metric criteria, ERCOT is considering a similar framework, and PJM, SPP, and MISO are considering using EUE in their planning criteria. Additionally, Public Service of Colorado and Tri-State Generation and Transmission Association are integrating a two-stage reliability criteria with stress testing. Globally, transitions like the Australian Energy Market Operator's adjustments to its EUE construct are also underway. In the European Union, the legislative framework establishes both LOLH and EUE as possible metrics for reliability standards. Currently, European Union member states typically define the reliability standard in terms of LOLH, and EUE is used as supplementary information.

While it remains viable for many regions to maintain a single criterion (LOLE or LOLH) in the near term, it is important to develop multi-metric criteria now, so that the resource adequacy framework is robust and adaptable as the resource mix evolves. This will allow planners to effectively meet future power system requirements and address emerging challenges. Delays will cause our resource adequacy framework to lag behind the rapid changes occurring to the resource mix, potentially leading to lurking reliability risks that are not captured in a single resource adequacy LOLE criterion.

This process could take place gradually over time, in two phases:

- **Phase 1:** Near-term resource procurement decisions (i.e., capacity market needs, planning reserve margin, and capacity accreditation) continue to be made using the existing single criterion, but report additional metrics and data visualizations (histograms and scatter plots of individual shortfalls) and apply stress-testing techniques for long-term planning.

- **Phase 2:** Additional criteria phase in over time, first including metrics like EUE and then including metrics that track outlier events (i.e., peak, energy, and duration VaR) and stress testing.

Shared Responsibilities

Implementing changes to the reliability standard in the power system requires a broad consensus among various stakeholders. The utilities and grid operators—in consultation with stakeholders—can lead this reform and be the ones to establish the resource adequacy framework, the analytical methods, and metrics used to measure adequacy. They are also responsible for devising specific plans or markets to meet these standards at reasonable costs.

However, the responsibility for determining the level, or minimum threshold, of the resource adequacy criteria ultimately falls on regulators—not the utility or power system planners. They play a crucial role in ensuring that the trade-off between risk and economic factors is appropriately balanced, as this is ultimately a societal and equity decision. This division of responsibilities ensures that while the regulatory bodies establish the criteria, the actual implementation is carried out effectively by those managing the power system.

The responsibility for determining the level, or minimum threshold, of the resource adequacy criteria ultimately falls on regulators—not the utility or power system planners. Regulators play a crucial role in ensuring that the trade-off between risk and economic factors is appropriately balanced, as this is ultimately a societal and equity decision.

Final Recommendations from the Task Force

This report acknowledges that no single solution fits all power systems regarding resource adequacy criteria. Expecting diversity in resource adequacy criteria is reasonable, and certain key considerations should be included in any new framework. Developed through broad industry consultation, the report identifies three essential components. These components, guiding planners and regulators in developing their frameworks, criteria, and resource adequacy thresholds, have been detailed throughout the report.

- **Transition to multi-metric criteria**
  - Loss-of-load expectation as the sole resource adequacy criterion represents only a single dimension of risk. It needs to be supplemented (p. 6).
  - Expected unserved energy is a preferred addition to incorporate size of shortfalls, especially as the system moves toward energy limitations (p. 15).
  - No one metric is the solution, and a multi-metric framework is needed to consider size, frequency, and duration of shortfalls (p. 33).
- **Specifically consider extreme events**
  - Not all resource adequacy loss-of-load events are the same. Tail risks can have a disproportionate impact on reliability and costs and should be quantified in resource adequacy criteria (p. 22).
  - Limited data are available to determine with confidence the probability of extreme events. This reality may require discrete analysis or stress-testing rather than a statistical measure (p. 29).
- **Incorporate economics**
  - The set of resource adequacy criteria should be used to establish the appropriate trade-off between reliability and cost. Cost and reliability are intrinsically linked, and this trade-off should be clear (p. 38).

There are multiple ways to accomplish these goals. To effectively navigate the energy transition, new resource adequacy criteria must encompass, at a minimum, a multi-metric approach, including both LOLE and EUE. Additionally, indicators should capture tail risks, and the framework must be more transparent about providing an economic justification for the chosen reliability level. This comprehensive approach, though challenging, is crucial for ensuring the reliability of our current, and evolving, power systems.

References

ACER (European Union Agency for the Cooperation of Energy Regulators). 2023. *Security of EU Electricity Supply*. Ljubljana, Slovenia. https://acer.europa.eu/sites/default/files/documents/Publications/Security_of_EU_electricity_supply_2023.pdf.

AEMC (Australian Energy Market Commission). 2023. *Review of the Form of the Reliability Standard and APC.* Sydney, Australia. https://www.aemc.gov.au/market-reviews-advice/review-form-reliability-standard-and-apc.

Billinton, R., and R. N. Allan. 1996. *Reliability Evaluation of Power Systems.* New York: Springer.

Calabrese, G. 1950. "Determination of Reserve Capacity by the Probability Method." *AIEE [American Institute of Electrical Engineers] Transactions* 50: 1681–1689.

Carden, K., N. Wintermantel, and J. Pfeifenberger. 2011. *The Economics of Resource Adequacy Planning: Why Reserve Margins Are Not Just About Keeping the Lights On.* Washington, DC: National Regulatory Research Institute. https://pubs.naruc.org/pub/FA865D94-FA0B-F4BA-67B3-436C4216F135.

Dent, C. 2019. "Capacity Markets, Risk Modelling and Decision Support." Presentation to the Isaac Newton Institute. https://sms.cam.ac.uk/media/2942328.

Dent, C., N. Sanchez, A. Shevni, J. Q. Smith, A. L. Wilson, and X. Yu. 2023. "Resource Adequacy and Capacity Procurement: Metrics and Decision Support Analysis." Pre-print. https://arxiv.org/pdf/2309.06119.pdf.

EPRI. 2022. "Resource Adequacy for a Decarbonized Future: A Summary of Existing and Proposed Resource Adequacy Metrics." White paper. Palo Alto, CA. https://www.epri.com/research/products/000000003002023230.

EPRI. 2023a. "Climate READi System Planning and Investment Prioritization." Q3 2023 Member Advisor and CRAG Webcast. Palo Alto, CA.

EPRI. 2023b. *Resource Adequacy for a Decarbonized Future, Case Study: Texas.* Palo Alto, CA. https://www.epri.com/research/products/000000003002027838.

ERCOT (Electric Reliability Council of Texas). 2023. "Reliability Standard Study Preliminary Results." Slide deck. Reliability and Markets Committee Meeting, June 19, 2023. Austin, TX. https://www.ercot.com/files/docs/2023/06/12/7-1-1-reliability-standard-study-preliminary-results.pdf.

ESIG (Energy Systems Integration Group). 2021. *Redefining Resource Adequacy for Modern Power Systems.* Reston, VA. https://www.esig.energy/resource-adequacy-for-modern-power-systems/.

ESIG (Energy Systems Integration Group). 2023. *Ensuring Efficient Reliability: New Design Principles for Capacity Accreditation.* Reston, VA. https://www.esig.energy/new-design-principles-for-capacity-accreditation.

EU (European Union). 2019. "Regulation (EU) 2019/943 on the Internal Market for Electricity." *Official Journal of the European Union.* The European Parliament and the Council of the European Union, June 5. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R0943&from=EN.

Fazio, J., and D. Hua. 2019. "Three Probabilistic Metrics for Adequacy Assessment of the Pacific Northwest Power System." *Electric Power Systems Research* 174: 105858. https://doi.org/10.1016/j.epsr.2019.04.036.

Felder, F. 2001. "An Island of Technicality in a Sea of Discretion: A Critique of Existing Electric Power Systems Reliability Analysis and Policy." *The Electricity Journal* 14(3): 21–31. https://doi.org/10.1016/S1040-6190(01)00183-X.

Gonzato, S., K. Bruninx, and E. Delaure. 2023. "The Effect of Short-Term Storage Operation on Resource Adequacy." *Sustainable Energy, Grids and Networks* 34: 101005. https://doi.org/10.1016/j.segan.2023.101005.

GridLab. 2022. *Reliably Reaching California's Clean Electricity Targets: Stress Testing Accelerated 2030 Clean Energy Portfolios.* Berkeley, CA. https://gridlab.org/wp-content/uploads/2022/05/GridLab_California-2030-Study-Technical-Report-5-9-22-Update1.pdf.

GridLab. 2023. *The Moonshot 100% Clean Electricity Study: Assessing the Tradeoffs Among Clean Portfolios with a PNM Case Study.* Berkeley, CA. https://gridlab.org/wp-content/uploads/2023/08/GridLab_Moonshot-Study_Report.pdf.

MISO (Midcontinent Independent System Operator). 2023. "Review 2024 Short-Term Incentive Plan Goals." Advisory Committee. September 13, 2023. Carmel, IN. https://cdn.misoenergy.org/20230913%20AC%20Item%2008%20Review%202024%20Short-Term%20Incentive%20Goals630175.pdf.

NERC (North American Electric Reliability Corporation). 2017. "Planning Resource Adequacy Analysis, Assessment and Documentation." Reliability standard BAL-502-RF-03. Atlanta, GA. https://www.nerc.com/pa/Stand/Reliability%20Standards/BAL-502-RF-03.pdf.

NERC (North American Electric Reliability Corporation). 2023. *2023 Long-Term Reliability Assessment.* Atlanta, GA. https://www.nerc.com/pa/RAPA/ra/Reliability%20Assessments%20DL/NERC_LTRA_2023.pdf.

NWPCC (Northwest Power and Conservation Council). 2024. "Discussion on Multi-metric Adequacy Metrics and Stakeholder Feedback." Resource Adequacy Advisory Committees. Portland, OR. https://www.nwcouncil.org/meeting/resource-adequacy-advisory-committees-2024-01-30/.

PJM Interconnection. 2023a. "Capacity Market Reform: PJM Proposal." July 27, 2023. https://www.pjm.com/-/media/committees-groups/cifp-ra/2023/20230727/20230727-item-02a---cifp---pjm-proposal-update---july-27.ashx.

PJM Interconnection. 2023b. “Capacity Market Reforms to Accommodate the Energy Transition While Maintaining Resource Adequacy.” Docket No. ER24-99-000. Federal Energy Regulatory Commission, December 1, 2023. https://www.pjm.com/-/media/documents/ferc/filings/2023/20231013-er24-98-000.ashx

PUCT (Public Utility Commission of Texas). 2023. “Reliability Standard for the ERCOT Market.” Project No. 54584. Austin, TX. https://www.puc.texas.gov/industry/electric/reliability.aspx.

SPP (Southwest Power Pool). 2024. “Supply Adequacy Working Group.” Zip file of meeting materials. February 20, 2024. Little Rock, AR. https://www.spp.org/Documents/71119/SAWG%20Meeting%20Materials%2020240220.zip.

Stenclik, D., G. Stephen, W. Cole, A. Tuohy, A. Bloom, and A. Figueroa Acevedo. 2022. “Beyond Expected Values: Evolving Metrics for Resource Adequacy Assessment.” Paris: CIGRE [International Council on Large Electric Systems]. https://cse.cigre.org/cse-n027/c5-beyond-expected-values-evolving-metrics-for-resource-adequacy-assessment.html.

Stephen, G., T. Joswig-Jones, S. Awara, and D. Kirschen. 2022. “Impact of Storage Dispatch Assumptions on Resource Adequacy and Capacity Credit.” 2022 17th International Conference on Probabilistic Methods Applied to Power Systems, June 12-15, 2022, Manchester, United Kingdom. New York: IEEE. https://ieeexplore.ieee.org/abstract/document/9810584.

Stephen, G., S. H. Tindemans, J. Fazio, C. Dent, A. Figueroa Acevedo, B. Bagen, A. Crawford, A. Klaube, D. Logan, and D. Burke. 2022. “Clarifying the Interpretation and Use of the LOLE Resource Adequacy Metric.” 2022 17th International Conference on Probabilistic Methods Applied to Power Systems, June 12-15, 2022, Manchester, United Kingdom. New York: IEEE. https://ieeexplore.ieee.org/document/9810615.

Tri-State Generation and Transmission Association. 2023. “2023 Energy Resource Plan.” Proceeding No. 20A-0528E. Westminster, CO. https://www.dora.state.co.us/pls/efi/efi_p2_v2_demo.show_document?p_dms_document_id=989882&p_session_id=.

UKDECC (United Kingdom Department of Energy and Climate Change). 2013. “Annex C: Reliability Standard Methodology.” London, UK. https://assets.publishing.service.gov.uk/media/5a7c4a16ed915d338141dd0a/Annex_C_-_reliability_standard_methodology.pdf.

PHOTOS

Cover: © iStockphoto/Mr. Socrates
p. viii: © iStockphoto/pkawasaki
p. xii: © iStockphoto/RossHelen
p. 1: © iStockphoto/zhongguo
p. 4: © iStockphoto/WichienTep
p. 7: © iStockphoto/Chris James
p. 10: © iStockphoto/helivideo
p. 13: © iStockphoto/Ngampol Thongsai
p. 16: © iStockphoto/Bulgac
p. 18: © iStockphoto/Daria Kulkova
p. 20: © iStockphoto/Petmal
p. 22: © iStockphoto/alexkich
p. 26: © iStockphoto/metamorworks
p. 29: © iStockphoto/urbanstar
p. 32: © iStockphoto/AndreyPopov
p. 33: © iStockphoto/wenpu wang
p. 36: © iStockphoto/sl-f
p. 40: © iStockphoto/Yozayo
p. 43: © iStockphoto/FatCamera

New Resource Adequacy Criteria for the Energy Transition: Modernizing Reliability Requirements

A Report by the Energy Systems Integration Group's Resource Adequacy Task Force

---

The report is available at https://www.esig.energy/new-resource-adequacy-criteria.

To learn more about ESIG's work on this topic, please send an email to info@esig.energy.

The Energy Systems Integration Group is a nonprofit organization that marshals the expertise of the electricity industry's technical community to support grid transformation and energy systems integration and operation. More information is available at https://www.esig.energy.

ESIG
ENERGY SYSTEMS
INTEGRATION GROUP