MEETING MATERIALS SUPPLEMENTAL PACKET

LIFE ACTUARIAL (A) TASK FORCE

March 22-23, 2025

NAIC SPRING NATIONAL MEETING

TABLE OF CONTENTS

<u>PAGE</u>

- 3 Agenda Item 2: VM-22 (A) Subgroup Meeting Minutes VM-22 Field Test Presentation
- 49 Agenda Item 4: GOES Presentation

Agenda Item 2

VM-22 (A) Subgroup Meeting Minutes VM-22

Field Test Presentation

Draft: 03/17/25

Valuation Manual (VM)-22 (A) Subgroup Virtual Meeting March 12, 2025

The VM-22 (A) Subgroup of the Life Actuarial (A) Task Force met Mar. 12, 2025. The following Subgroup members participated: Ben Slutsker, Chair (MN); Elaine Lam, Vice Chair (CA); Lei Rao-Knight (CT); Mike Yanacheak (IA); Matt Cheung (IL); William Leung (MO); Seong-min Eom (NJ); William B. Carmello (NY); Rachel Hemphill and Iris Huang (TX); Tomasz Serbinowski (UT) and Craig Chupp (VA).

1. Adopted standard projection amount as a disclosure requirement

Slutsker said the standard projection amount (SPA) calculation worked like the normal stochastic calculation but instead of using company assumptions, the SPA calculation used prescribed assumptions. If the CTE 70 calculated from the SPA calculations was higher than the stochastic reserve, then that CTE 70 based on prescribed assumptions would go through the buffer calculation. He stated that the decision can be revisited in a couple years regardless of the decision to make the VM-22 SPA a binding reserve floor or a disclosure only item for implementation. Slutsker noted that the SPA floor rarely kicked in for VM-21, Requirements for Principle-Based Reserves for Variable Annuities.

Angela McShane (EY) provided a summary of the SPA and CTE 70 from VM-22 field test results. EY performed a comparison of the stochastic reserve versus the unbuffered SPA and buffered SPA. She said for the most part, the SPA was usually less than the stochastic reserve when looking at the unbuffered amount. Leung asked for the magnitude of the additional amount when the SPA is larger and recommended reviewing the field test results in a regulator only session. Steve Jackson (American Academy of Actuaries—Academy) said by looking at the range of results presented a sense of outliers can be interpreted. Jackson said the range between the SPA and the stochastic reserve was larger for the fixed index annuity (FIA) products.

Cheung asked if the SPA assumptions were set at a moderately adverse level or average. Lam said that there were no explicit margins on the SPA assumptions and were intended to represent the average. Eom said the same applies to mortality, there were no explicit margins and would be considered average. Slutsker said the buffer was designed to catch more of the outliers by using average assumptions as opposed to moderately adverse.

Brian Bayerle (American Council of Life Insurers—ACLI) said the ACLI recommends the SPA be a nonbinding disclosure item. He said the SPA fails to reflect the diversity of products in the scope of VM-22 to be as effective at capturing outlier assumptions as the assumptions are currently designed. The disclosure requirement would be sufficient to meet the goal of the SPA by allowing regulators to review the VM-31 PBR Actuarial Reports to identify outlier companies with SPAs greater than the CTE 70. The ACLI noted concerns that a binding SPA could undermine ongoing efforts with the generator of economic scenarios project intention to incentivize hedging as well as the reinsurance asset adequacy testing project concerning the types of products in scope of VM-22.

Cheung said the VM-31, PBR Actuarial Report Requirements for Business Subject to a Principle-Based Valuation may need to be enhanced to collect detailed information given that the existing disclosure requirements were based on assuming an SPA floor. Slutsker said the VM-31 requirements were edited in 2024 to create a single annuities section encompassing VM-21 and VM-22 and there is a section on the SPA. Slutsker said the Subgroup may need to revisit the VM-31 edits if the SPA is a disclosure only decision. Slutsker also said the proposed supplement to the annual statement collects data that could serve as a template for the information that could

help form a repository based on annual statement information. Chueng said since there will be the ability for companies to use their best estimate assumptions as opposed to prescribed, then there should be mandatory disclosure items if the route is disclosure only. Bayerle said the ACLI would work with the Subgroup to develop language that would get regulators what they need to understand the appropriateness of the assumptions and why it was appropriate for the company not to post a binding reserve with respect to the SPA.

Bruce Friedland (Academy) said the Academy's position in 2023 when the issue was first raised was that the SPA should be a disclosure only because it better reflected the spirit of PBR. He said the net premium reserve (NPR) binding floor under VM-20 and the SPA floor under VM-21 were implemented for different reasons. The NPR was introduced partially for tax purposes and was not initially meant to be part of the principle-based framework and was not implemented for the same reasons as the VM-21 floor. Friedland said for VM-22, an SPA disclosure provides the information needed to identify outliers. He noted the field test results did not have enough information to decide because if the SPA was higher, it did not necessarily mean the underlying assumptions for the stochastic reserve were unreasonable or inappropriate. The Academy supported the SPA being a disclosure item at least until there was more information available.

Andy King (Oliver Wyman) said Oliver Wyman performed model office testing and waterfall attribution going from their scenario reserve to the prescribed projection amount. He said across the products they tested the SPA was generally not binding and consistent with the VM-22 field test results. The main drivers were the lapse and mortality assumptions that resulted in the SPA being 6-7% lower than the stochastic reserves. He said it was due to their more conservative assumptions for the withdrawal benefit block where they tried to use industry average assumptions that companies would typically assume for mortality and lapse. King said he could share the results with the Subgroup members.

Yanacheak said he did not recommend the SPA as a floor because it did not allow for the proper variation in the types of products represented. The incentives that may cause a policyholder to behave in a particular way will vary from product to product even within the same category, such as with living benefit designs, and can result in meaningless average behavior assumptions. Companies may have credible data that show an average assumption is too conservative or too aggressive. He said the situation where there is not credible experience is where the industry average used for the SPA caused him concern with relying on a floor. Regulators need to understand the risks a company is taking and the process used to arrive at the reserve assumptions. He said an SPA floor may create a false sense of security and shift attention from that understanding. Cheung said from his perspective, it would go against the spirit of PBR to force companies to use a different assumption just because the industry average was different than their fully credible assumption that can be justifiably different than the average.

Lam said a floor could come into play or be beneficial when companies have low credibility. She said she was inclined to maintain consistency with VM-20 and VM-21 due to lack of full information and experience of how VM-22 will work in practice. Lam noted that regulators should review those assumptions very carefully and the SPA as a floor would serve as an additional backstop that would not absolve regulators of the responsibility to review the stochastic reserve calculation. Rao-Knight suggested collecting additional data during the 3-year implementation period where some companies could try to go through the exercise so that more experience data could be collected. Cebula said New York also preferred a floor.

Eom said the *Valuation Manual* amendment proposal to collect group annuity experience could help with some of the lack of data. She agreed that the wide array of the products and the average may not represent the products for which the SPA is calculated.

The Subgroup agreed to move forward with the SPA as a disclosure item rather than a binding floor.

Having no further business, the VM-22 Subgroup adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/A CMTE/LATF/2025-1-Spring/VM-22 Calls/03 12/Mar 12 VM22Minutes.docx

Draft: 03/18/25

Valuation Manual (VM)-22 (A) Subgroup Virtual Meeting March 5, 2025

The VM-22 (A) Subgroup of the Life Actuarial (A) Task Force met Mar. 5, 2025. The following Subgroup members participated: Ben Slutsker, Chair (MN); Elaine Lam, Vice Chair (CA); Lei Rao-Knight (CT); Mike Yanacheak (IA); Matt Cheung (IL); William Leung (MO); Seong-min Eom (NJ); Bill Carmello (NY); Rachel Hemphill and Iris Huang (TX) and Tomasz Serbinowski (UT).

1. <u>Received a Status Report on Closed Meeting</u>

The Valuation Manual (VM)-22 (A) Subgroup met in regulator-to-regulator session Mar. 3rd, 2025, pursuant to paragraph 3 (specific companies, entities or individuals) of the NAIC Policy Statement on Open Meetings. During the meeting, the Subgroup took no action.

2. Adopted Six Percent Stochastic Exclusion Ratio Test Threshold

Slutsker said the regulators recommended 6% for the SERT threshold based on their review of company results from the VM-22 field test, which also aligned with the SERT threshold under VM-20, Requirements for Principle-Based Reserves for Life Products. Slutsker said for fixed deferred annuities (FDAs) without guarantees, the Subgroup was comfortable with those products passing the SERT. He said in general, fixed deferred annuities with guarantees and payout annuities would generally not be able to pass the SERT. The 6% threshold was conservative enough to provide room for FDAs and fixed indexed annuities (FIAs) without guarantees to pass while low enough to scope in many of the other contracts like single premium immediate annuities (SPIAs), structured settlements and contracts with living benefits. Slutsker said the SERT would be less important in the first three years because a company could choose not to implement it until the end of the three-year optional implementation period. Slutsker noted that in addition to the SERT as an exclusion test under VM-22, there is also a certification test option that requires documentation of any rationale supporting the test as part of VM-31, PBR Actuarial Report Requirements for Business Subject to a Principle-Based Valuation.

Bruce Friedland (American Academy of Actuaries - Academy) said that they did not have information to provide recommendations, so they suggested using a placeholder threshold such as the 6% from VM-20. The Academy offered three other options: 1) using any updated threshold following a formal generator of economic scenario (GOES) adoption; 2) performing additional model office testing in cooperation with EY; and 3) following up with field test participants to get more information.

Brian Bayerle (American Council of Life Insurers - ACLI) stated the ACLI could not make a specific threshold recommendation due to lack of publicly available data. The ACLI provided principles to consider the threshold selection: 1) it should result in negligible variation of products passing in and out of the SERT from one year to the next when the risk associated with those products has not changed; 2) it should not be selected from one model office or one company's data; and 3) modifications to the deterministic certification option may be warranted because there are going to be situations where a stochastic reserve is not going to be adding value over a deterministic reserve when there is not a true risk of volatility associated with the interest rate.

Leung, seconded by Rao-Knight, made a motion to reflect 6% as the SERT threshold in the VM22 framework. The motion passed with New York abstaining.

3. Adopted One-Percent Mortality Sensitivity for SERT

Slutsker said the purpose of the mortality sensitivity is to try and capture that risk factor. He said pre-PBR CARVM for payout annuities used prescribed mortality tables instead of the company's own mortality with additional margin. The prescribed mortality table under CARVM may not be appropriate for a block of payout annuities and that was part of the rationale to boost the mortality sensitivity from the 5% mortality factor in the current VM-22 draft.

Friedland (Academy) said the Academy recommended keeping the 5% mortality factor for the sensitivity because the impact was not significant. The Academy noted a 5% mortality adjustment was a reasonable number because the SERT was a measure of interest and asset risk and a change in mortality should not be the sole driver for a product being subject to a stochastic reserve. The Academy also noted in their comment letter several drivers that could change the magnitude of the 5% mortality sensitivity observed in model office results: 1) the relatively high average attained age for the model office field test block lowered the SERT result and its sensitivity to mortality shocks; 2) only life with certain for SPIA were modeled and that lowered the mortality sensitivity SERT result; and 3) with VM-22 being prospective only, the sensitivity is likely to be more impactful on a single-issue year than a full block.

Slutsker said the proposal to reflect a 1% mortality improvement sensitivity would only be applied as future mortality improvement (FMI) which meant it will be the percentage of reduction compounded each year from the valuation date to that given projection year within the calculation. It does not refer to historical mortality improvement. Slutsker said with respect to FMI there is uncertainty with a trend variable. There are different opinions between experts in the field of whether the FMI trend could be more mortality improvement or deterioration. Some of the rationale for changing to use FMI was that there is uncertainty with how future technology, different diseases, and the rate of medical advancements will impact mortality.

Slutsker summarized field test results shared publicly on the Subgroup's Feb. 5 meeting. The results showed a small impact of at most 1% under the 5% mortality factor sensitivity and even the 10% mortality factor sensitivity was not impactful. Angela McShane (Ernst Young – EY) said EY modeled fixed annuities without guaranteed withdrawal benefits and there was an immaterial impact from a change to using mortality improvement. She said for fixed annuities with withdrawal benefits the SERT ratio increased a little from around 2.2% to around 3% by changing the sensitivity from a 5% scalar to 1% mortality improvement. McShane said the results from other products were still under review.

Lam, seconded by Eom, made a motion to use the 1% future mortality improvement sensitivity in the VM-22 framework. The motion passed. New York abstained.

4. Discussed Exclusion Testing for Future Hedging Programs Supporting Indexed Credits

Slutsker said during the regulator only call a question came up regarding whether hedging programs for indexed credits should be allowed to test for exclusion. He said the Subgroup at one point was comfortable allowing fixed indexed annuities to be eligible for the exclusion test. Slutsker discussed an example that under VM-20 an indexed universal life product with a future hedging program would not be eligible for the exclusion test. Hemphill noted that allowing some exclusion testing would be a significant departure from VM-20. It would require additional VM-31 disclosures for documentation on hedging and a clear outline of what type of hedging would be eligible for exclusion testing. Hemphill said a new section would need to be drafted and she was hesitant to add another area of work for the project given the timeline for implementation. Slutsker said the Subgroup could plan to allow for it and work on drafting the SERT deviation during the phase in period. Cheung said he does not have a strong preference, but if the Subgroup decided to allow some of these products to be eligible to pass the SERT, then the

eligibility scope should be narrow. Cheung noted that it may be difficult to define what would be eligible and a decision to disallow these index annuities to pass the exclusion test would be simplest from a documentation and drafting standpoint. Slutsker said the Subgroup will revisit this topic on a future call.

5. Discussed Other Matters

Hemphill said regarding the SERT, the VM-22 draft contained the current VM-20 language. She said there has been discussion of the need to change from using anticipated experience to prudent estimate experience. She said for the SERT there is no hedge breakage, but when the actual modeling run is done there would be a hedge breakage assumption. She said this is another area that highlights the disconnect between doing the SERT on an anticipated basis whereas the stochastic reserve is on a prudent basis. Hemphill said she included in the GOES draft a consideration to include margins because it relates to the idea of revisiting the SERT threshold. She said she also proposed in the GOES draft to expand the documentation in VM-31 as well as be consistent with what is measured using prudent assumptions under VM-20. Slutsker said the decisions on GOES would go into VM-22 as appropriate.

Having no further business, the VM-22 Subgroup adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/A CMTE/LATF/2025-1-Spring/VM-22 Calls/03 05/Mar 5 VM22Minutes.docx

Draft: 03/18/25

Valuation Manual (VM)-22 (A) Subgroup Virtual Meeting February 26, 2025

The VM-22 (A) Subgroup of the Life Actuarial (A) Task Force met Feb. 26, 2025. The following Subgroup members participated: Ben Slutsker, Chair (MN); Elaine Lam, Vice Chair (CA); Lei Rao-Knight (CT); Mike Yanacheak (IA); Matthew Cheung (IL); William Leung (MO); Seong-min Eom (NJ); Bill Carmello (NY); Rachel Hemphill and Iris Huang (TX) and Tomasz Serbinowski (UT).

1. Adopted the Texas reinvestment guardrail proposal for VM-22

Slutsker said the Subgroup was awaiting field test results to discuss the reinvestment guardrail under the proposed VM-22, Requirements for Principle-Based Reserves for Non-Variable Annuities and how it may align or differ from VM-20, Requirements for Principle-Based Reserves for Life Products and VM-21, Requirements for Principle-Based Reserves for Variable Annuities. The three reinvestment guardrail options consist of fixed income assets that have the same weighted average life as the company modeled strategy that are all public non-callable corporate bonds with gross asset spreads, asset default costs, and investment expenses by projection year that are consistent with a credit quality blend of: a) 50% AA and 50% A from the current VM-20/VM-21 reinvestment guardrail; b) 5% Treasury, 15% AA, 40% A, and 40% BBB as proposed by the Academy; or c) 5% Treasury, 15% AA, and 80% A as proposed by Texas. Slutsker said the VM-22 field test model office results presented during the 2024 Life Actuarial Task Force (LATF) Fall National Meeting indicated the impacts between the guardrail options were small. He said the VM-22 field test participant results presented at the Feb. 5 Subgroup call were also small. Carmello said that with surplus ratios as low as they are, a small impact could still have a significant impact on surplus. Angela McShane (EY) said that the C-3 Phase 1 results from the VM-22 field test were not ready to present yet.

Serbinowski said that regardless of which guardrail option the Subgroup adopted, there was no expectation that companies would reinvest that way. He said the guardrail decision would not mandate how the company could reinvest but it would mandate what could be used for the valuation. Hemphill noted that if the reserves were higher under the company's actual reinvestment strategy, the company should use their company strategy rather than the guardrail since using the guardrail is intended to work as a minimum. Hemphill stated a company would always compare the guardrail to their actual investment strategy.

Bruce Friedland (American Academy of Actuaries—Academy) said the Academy still recommended the Academy proposal because it was more consistent with investment practices among the companies while maintaining conservatism by reflecting investment-grade quality. Link Richardson said fixed income assets like commercial mortgages, structured securities, and private placements tend to have higher yields and were not assumed in the guardrail mix and would provide another element of conservatism. Hemphill said equities were not included because the guardrail was composed only of fixed income assets. Carmello said that a conservative approach was justified due to the uncertainty associated with future investments that have yet to be made. Hemphill agreed with Carmello.

Brian Bayerle (American Council of Life Insurers—ACLI) said ACLI preferred the Academy proposal because it better aligned with company reinvestment strategies and the products and guarantees it supported. He indicated the guardrail would likely be binding under VM-22 since it had been binding under VM-20 and VM-21. Bayerle said the Texas proposal may be more appropriate for VM-20 and VM-21. ACLI also supported the Texas proposal as a compromise.

Cheung asked what impact the Subgroup's decision for VM-22 would have on VM-20 and VM-21. Hemphill noted that under the VM-22 project, the Subgroup has occasionally made independent decisions. She emphasized that consistency should be considered when updating *Valuation Manual* chapters unless there is a valid reason not to. She said sometimes new knowledge becomes available that may justify a deviation from the other chapters. She added that if the Subgroup picked something different, then the Subgroup should make a referral to LATF. Hemphill said she did not see a good reason for VM-20, VM-21, and VM-22 to deviate but it can be discussed during a LATF proposal. Yanacheak said a project to investigate the guardrail under all three chapters was worthwhile, however he was unable to make an informed decision that would impact VM-20 or VM-21. Scott O'Neal (NAIC) said NAIC has models for products under VM-20 and VM-21 that could be used to analyze the reinvestment guardrail impact. Slutsker clarified that the decision from the Subgroup should come from appropriateness with respect to VM-22, and LATF would address appropriateness for VM-20 and VM-21. Richardson said if the Subgroup decided to refer to LATF it should also make a referral to the Life Risk-Based Capital (E) Working Group because of C-3 testing.

Yanacheak said he was uncomfortable with the idea that changing the reinvestment strategy could reduce the liability. He recommended going with the reinvestment guardrail currently in VM-20 and VM-21 because the results for the options show small impacts. He stated that LATF could then do an investigation into a change to the guardrail because the data did not appear to support a deviation. Carmello said he also preferred using the same reinvestment guardrail currently used in VM-20 and VM-21.

Lam said California supports the Texas proposal because it fell between the other two proposals. Hemphill said if the Subgroup thinks about the guardrail in isolation for VM-22, the difference does not appear to matter. She said she had a concern that the current guardrail could be overly constraining and preferred the Texas proposal that fell in the middle between the current VM-20/VM-21 guardrail and the Academy proposal. Slutsker noted that Craig Chupp (Virginia) was unable to attend the call and indicated a preference for the Texas option.

Cheung asked if the current 50% AA and 50% A guardrail was arbitrary or was there significant data to justify the assumption. He said he was concerned if the Subgroup decided to deviate from the guardrail in other chapters that they could be moving away from a well thought out assumption. Slutsker said he was not part of the initial conversation on the VM-20 guardrail development and was unaware of a survey of assets. However, for the rates under the current VM-22, Statutory Maximum Valuation Interest Rates for Income Annuities, industry data was reviewed by the NAIC. Slutsker said the intent for VM-20 was more about conservatism to limit additional spread that could get collected on top of the risk-free rate for discounting so as not to incentivize certain assumptions of reinvestment that result in higher discount rates and lower reserves.

Slutsker said a difference between cash flow testing and principle-based reserving (PBR) is that you get prescribed defaults and reinvestment spreads in PBR. Under VM-20 and VM-21 there is a VM-31 report disclosure that shows the impact between the alternative reinvestment strategy and the company's own reinvestment assumptions. Slutsker said when he reviewed various VM-31 PBR Actuarial reports the percentage impacts aligned with the VM-22 field test results, though the significance may depend on the surplus level.

The Subgroup voted to use the Texas proposal of 5% Treasuries, 15% AA and 80% A for the VM-22 reinvestment guardrail.

Having no further business, the VM-22 (A) Subgroup adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/A CMTE/LATF/2025-1-Spring/VM-22 Calls/02 25/Feb 25 VM22Minutes.docx

Draft: 3/14/25

Valuation Manual (VM)-22 (A) Subgroup Virtual Meeting February 19, 2025

The VM-22 (A) Subgroup of the Life Actuarial (A) Task Force met Feb. 19, 2025. The following Subgroup members participated: Ben Slutsker, Chair (MN); Elaine Lam, Vice Chair (CA); Lei Rao-Knight (CT); Mike Yanacheak (IA); Matt Cheung (IL); William Leung (MO); Seong-min Eom (NJ); Bill Carmello (NY); Rachel Hemphill and Iris Huang (TX) and Craig Chupp (VA).

1. Discussed Comments Received on SPA Policyholder Behavior Assumptions

A. Partial Withdrawal Assumption by Distribution Channel

Brian Bayerle (American Council of Life Insurers—ACLI) suggested that partial withdrawal assumptions vary by distribution channel. Lam said that while the withdrawal data is available by distribution channel, it would be difficult due to data availability and confidentiality. She said the question would become which assumptions could be split. Cheung noted that as long as the requirement to assess the impact of aggregation is on the additional standard projection amount (SPA) then the differentiation by distribution channel may be implicitly captured. Lam recommended retaining the current level of assumption breakdown, which does not vary by distribution channel.

B. <u>Withdrawal Commencement for Accumulation Reserving Category with GLB—Section 6.C.4.c</u>

Bruce Friedland (American Academy of Actuaries—Academy) recommended companies use their own best estimate assumption rather than having a forced full utilization by a certain point in time as indicated in Section 6.C.4.c because it better meets the goal of the SPA to catch outliers. Jonah von der Embse (Academy) said that a prescribed utilization assumption, as written in the prior draft, may not be conservative for some product designs.

Lam said the reason the assumption was changed to a prescribed assumption without a comparison was because commenters indicated it was difficult to compare their best estimate to the prescribed at different projection years. Lam said her new recommendation is for companies to use their best estimate but with a guardrail to ensure that a specified percentage of the contracts are withdrawing by a certain age or contract year. Lam said this recommendation is a compromise that allows flexibility to accommodate different product designs where companies may have more conservative assumptions as well as simplify implementation.

Slutsker said this approach would work like using the guardrail as replacement of assumptions rather than a comparison. Friedland (self) said he also interpreted the edits to be used as an assumption rather than a comparison or a floor by replacing the assumptions to make sure it works to meet the goals of the SPA. Bayerle said ACLI has concerns that using a company's best estimate creates an inconsistency in the framework. Bayerle said the ACLI will consider this approach and provide feedback.

Friedland (Academy) said the structure is generally consistent with the Academy proposal, but some of the contract year values in the draft were different than the last exposure in Section 6.C.4.c.i-ii. Lam stated that she updated the percentages of contracts and the contract years to align more closely with the data the VM-22 Policyholder Behavior drafting group received from LIMRA.

C. Dynamic Lapse Formula for Full Surrenders—Section 6.C.5

Lam said the dynamic lapse formula contains an in-the-moneyness (ITM) component, and there are lapse tables for guaranteed living benefits (GLBs) that also vary by ITM. Lam said these lapse rates were intended to be base lapse rates to apply to the policies and then it goes through the dynamic lapse formula. Lam said the ACLI raised a concern that the base lapse rates in the tables and the dynamic lapse formula both account for ITM. She said the ACLI also questioned the high level of the lapse rates looks incorrect.

Lam said after doing some research, she does agree the moneyness may be accounted for twice. Lam said one of the reasons the numbers may seem a little high and not be intuitive is related to the level of granularity that the data is cut so some of the cells may not have sufficient credibility. Lam said one solution may be to simplify the base lapse assumptions by eliminating the split by ITM, retaining the split by age and years-to-expiry, and addressing the ITM in the dynamic lapse formula. Slutsker said that the assumptions to use for contracts prior to utilization and after utilization may be different. Slutsker said the issue of double counting the moneyness may be resolved by answering the question of whether the dynamic lapse assumption was built on the base lapse or built independently. Lam said the drafting group will address these comments by performing additional analysis of the data.

Bayerle said the ACLI agreed with the concept that the market-value-adjustment (MVA) factor should be zero when the MVA is in effect. However, because the rate factor is additive when the MVA is not in effect when the difference between the market and crediting rates is significant, it could create a situation where the impact for the market rate would be smaller when the base lapse rates are higher and larger impacts when the base lapse rate is small. Bayerle said it may be easier to adjust the cash surrender value (CSV) based on the lapse function than the current MVA. Lam said part of the challenge with using a multiplicative factor on the base rates is that it multiplies everything else. Lam said she will perform additional reviews of the data and examples to see if it is a major issue and if there is a way to alleviate the impact. Bayerle said the ACLI will also brainstorm some ideas and see if they can come to a consensus.

Bayerle said, as structured, the lapse rates coming out of the dynamic lapse formula have a high degree of sensitivity to interest rates, and the only limit is the 90% cap. He said the 90% may work more generally, but one way to address the issue is to make the maximum lapse rate vary by product type, such as those with and without GLBs. Lam said making it more granular could minimize or restrict any one of the components, which is not the desired intent. She said it may be more efficient to review the market factor or other factors in the formula and adjust those rather than increasing the granularity of the minimums and maximums. Bayerle said ACLI would take it back to members for input.

Having no further business, the VM-22 Subgroup adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/A CMTE/LATF/2025-1-Spring/VM-22 Calls/02 19/Feb 19 VM22Minutes.docx

Draft: 03/09/25

Valuation Manual (VM)-22 (A) Subgroup Virtual Meeting February 12, 2025

The VM-22 (A) Subgroup of the Life Actuarial (A) Task Force met Feb. 12, 2025. The following Subgroup members participated: Ben Slutsker, Chair (MN); Elaine Lam, Vice Chair (CA); Lei Rao-Knight (CT); Mike Yanacheak (IA); Seong-min Eom (NJ); William B. Carmello (NY); and Rachel Hemphill and Iris Huang (TX).

1. Discussed Reserve Floor Options and Allocation Methodology for Longevity Reinsurance Transactions

Slutsker said that the Subgroup decided during its Dec. 11, 2024, meeting to use the 2% of benefits floor approach proposed by New Jersey but did not decide when to apply the floor. The options proposed to the Subgroup were to apply the floor at a contract level, scenario level, or in aggregate.

Brian Bayerle (American Council of Life Insurers—ACLI) discussed the ACLI approach to apply the floor to the final reserve after calculating the conditional tail expectation (CTE). He said the ACLI recommends its approach because it is simple to implement, is consistent with the allocation methodology, and complies with the Subgroup's consensus to floor longevity reinsurance reserves above zero in early durations. Bayerle said since the flooring calculation can be done outside the principle-based reserving (PBR) model it is less prone to errors and would be easier to validate. He said the ACLI does not expect the benefits to vary between scenarios because of a lack of stochastically modeling mortality and the lack of non-U.S. dollar exchange and interest rates in the NAIC/Conning generator of economic scenarios (GOES). Bayerle added that even if companies voluntarily stochastically modeled mortality or accounted for the currency issue, the variability of the benefits across scenarios would be minimal in the first 12 months since that would be expected to appear much later in the projection. Bayerle said the ACLI does not recommend applying the floor at a policy level because it would require stochastic and asset calculations for each individual policy that is inconsistent with the rest of the stochastic calculations.

Eom said her proposal to apply a 2% floor at the scenario reserve level prior to the CTE70 calculation avoids low and negative reserve amounts resulting in higher CTE70. She noted that the ACLI approach to apply the 2% in aggregate results in a lower CTE70 because the floor applies after the averaging which allowed the negative scenario reserves to impact the average. Eom said she supported the ACLI's approach regarding the allocation methodology. Hemphill agreed with Eom that the scenario reserves level was the appropriate place to floor. Hemphill said that the approach was consistent with the cash surrender value floor under VM-21, Requirements for Principle-Based Reserves for Variable Annuities, where the flooring occurs at the scenario level.

The Subgroup agreed to move forward with the 2% of annual longevity benefits floor at the scenario reserve level proposed by New Jersey and the ACLI's allocation methodology proposal with a change to the proposal to make it so there is a floor within the allocation methodology at the 2% of annual longevity benefits.

2. Discussed Other Matters

Slutsker noted that during a future meeting, the Subgroup will discuss the payout annuity threshold exclusion test. He asked companies and other interested parties to submit insights they have regarding the threshold level.

Having no further business, the VM-22 (A) Subgroup adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/A CMTE/LATF/2025-1-Spring/VM-22 Calls/02 12/Feb 12_VM22Minutes.docx

© 2025 National Association of Insurance Commissioners 1

Draft: 2/24/25

Valuation Manual (VM)-22 (A) Subgroup Virtual Meeting February 5, 2025

The VM-22 (A) Subgroup of the Life Actuarial (A) Task Force met Feb. 5, 2025. The following Subgroup members participated: Ben Slutsker, Chair (MN); Elaine Lam, Vice Chair (CA); Lei Rao-Knight (CT); Mike Yanacheak (IA); Matt Cheung (IL); William Leung (MO); Seong-min Eom (NJ); William B. Carmello (NY); Rachel Hemphill and Iris Huang (TX); Tomasz Serbinowski (UT); and Craig Chupp (VA).

1. Heard a Presentation on VM-22 Field Test Results

Angela McShane (EY), Kyle Stolarz (EY), and Steven Jackson (American Academy of Actuaries—Academy) presented the preliminary summary of the VM-22 field test results to the Subgroup (Attachment A). McShane noted a key area that companies commented on was simplifications or limitations in their results regarding the assets. She said that in general, the field test results aligned well with the model office results presented at the 2024 Fall National Meeting. Stolarz presented the available field test sensitivity results noting that to maintain confidentiality, the Stochastic Exclusion Ratio Test (SERT) results could not be presented publicly.

Slutsker stated the field test results showing reserve increases for contracts with living benefits seemed counterintuitive because contracts with optionality were expected to see decreases since the Commissioners Annuity Reserve Valuation Method (CARVM) takes the highest of all scenarios. McShane said that companies that saw an increase in reserves noted they had limitations and made simplifications regarding the asset portfolio. McShane said those participants indicated that they would be doing additional analysis and work to optimize their asset portfolio for VM-22.

Slutsker said the sensitivities results were key to future discussions on making decisions on open topics. McShane said companies indicated that reinvestments had a large impact when the reinvestment guardrail was compared to the company's own distribution, but the impact was small when comparing the baseline results to the guardrail sensitivity results. McShane stated that while the SERT results could not be shared to preserve anonymity, generally the results were consistent with the model office results.

2. Exposed Outstanding Decisions for VM-22 Draft Requirements

Slutsker exposed open questions on topics regarding the reinvestment guardrail, SERT threshold, and SERT mortality sensitivity for a 21-day public comment period ending Feb. 25. He also exposed a request for any other revisions to the framework for a 40-day public comment period ending Mar. 17 (Attachment B).

Having no further business, the VM-22 (A) Subgroup adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/A CMTE/LATF/2025-1-Spring/VM-22 Calls/02 05/Feb 5_VM22Minutes.docx

Draft: 3/9/25

Valuation Manual (VM)-22 (A) Subgroup Virtual Meeting December 11, 2024

The VM-22 (A) Subgroup of the Life Actuarial (A) Task Force met Dec. 11, 2024. The following Subgroup members participated: Ben Slutsker, Chair (MN); Elaine Lam, Vice Chair (CA); Lei Rao-Knight (CT); Matt Cheung (IL); Nicole Boyd (KS); Seong-min Eom (NJ); William B. Carmello (NY); Rachel Hemphill and Iris Huang (TX); Tomasz Serbinowski (UT); and Craig Chupp (VA).

1. Discussed Longevity Reinsurance Transactions Reserve Floor Methodologies

Slutsker said that longevity reinsurance transaction (LRT) products with recurring premiums could, on a presentvalue basis, result in negative reserves. He said to prevent negative reserves from offsetting any other positive reserves from other products, the Subgroup created a separate reserving category for LRT so that it could not be aggregated with any other types of contracts. He said the other way the VM-22, Requirements for Principle-Based Reserves, draft addressed potential negative reserves was with a k-factor method. Slutsker said the k-factor methodology proposed in the VM-22 draft worked similarly to a net-level premium method by setting the k-factor so that the present value of benefits equal to the present value of premiums at time zero, and the k-factor would then be applied to premiums going forward for stochastic reserve calculations. Slutsker said two proposed alternatives to the k-factor methodology are a percentage of benefit floors. The methodology proposed by the American Council of Life Insurers (ACLI) floored the reserve at zero. He said the other methodology proposed by New Jersey is a 2% annual benefits floor methodology similar to the ACLI proposal.

Brian Bayerle (ACLI) said the original ACLI proposal had a zero floor after calculating the conditional tail expectation 70 (CTE 70) for each individual contract. He said a 2% proposal is a very different floor than zero, and they did not envision the application at the scenario level. Bayerle recommended modifying the ACLI proposal so that there would be no per-contract floor, which was different than its original proposal and New Jersey's proposal.

Eom said that the 2% method New Jersey proposed would apply on a scenario basis when companies project cash flows before calculating the CTE 70 of the scenario reserves. The reason for the 2% floor at that level was to address the benefit amount variation across scenarios when companies use stochastic mortality assumptions or mortality experience assumptions from foreign countries. Eom noted that the Subgroup would need to discuss how to allocate back to the treaty level.

Mark Hutchinson (American Academy of Actuaries—Academy) discussed how LRT contracts work and how negative reserves may happen (Attachment A). He said the Academy viewed reserve floors as inconsistent with a principles-based framework. He said the Academy was not opposed to the 2% floor approach; however, for an actuary setting the assumptions, a 0% floor would suffice. Hutchinson said the VM-31/VM-G requirements would provide transparency to the assumption-setting process to allow regulators to gain comfort with a 0% floor.

Cheung asked if there would always be a standard projection amount (SPA) for LRT. Slutsker said that there is some similarity to pension risk transfer (PRT) where there could be some variability in how the mortality profile compares to the mortality under the SPA, so there may not necessarily always be an SPA. Hutchinson said that these agreements are often bespoke to the specific group that will be reinsured, and there are different longevity expectations, so the SPA may or may not kick in. Cheung asked if there were essentially two floors, a floor discussed today, and the mortality that may act as a floor. Slutsker said regarding the "double floor," there has

© 2024 National Association of Insurance Commissioners 1

been no formal decision on whether to recommend that the SPA be a floor or a disclosure item, but it will be decided in the coming months. Cheung said it would be difficult to agree on a decision if it is unknown whether the SPA will be required as a floor or disclosure. Slutsker said there will be an opportunity before adoption to discuss various questions and decisions.

Bayerle said ACLI prefers not to have a floor because it does not fully align with the economics, but they understand regulators' concerns about early profits from the LRT products. He said ACLI would be comfortable if regulators were to move forward with a 2% floor approach as it is a simplified approach and balances setting reserves earlier.

Eom said she preferred the 2% approach and would like to see testing of the different options for applying the floor. Lam said California supported the 2% floor proposal and wanted to better understand the impact of how it would apply at the different points in the calculation. Leung, Rao-Knight, Boyd, and Cheung said they supported the 2% floor as proposed by New Jersey. Carmello said New York favored the k-factor approach.

Eom made a motion, seconded by Leung, to incorporate the 2% benefit floor approach into the draft with the understanding that: 1) there are three places where the floor could be applied; and 2) the Subgroup's decision of the SPA being a floor or disclosure will come later. Cheung said Illinois supports the 2% but will preserve his final decision to where the decision is made for the SPA to be a disclosure or floor. The motion passed with New York opposed.

2. Discussed Applications and Timing of LRT Reserve Floor Methodology

Slutsker said the 2% floor could apply independently for each contract or in aggregate and at different points in the calculation such as at the scenario reserve level before the CTE calculation or during the stochastic reserve calculation after the CTE.

Eom said the scenario reserve level calculation would work so that for each scenario, the floored scenario reserve would be set to the maximum of the scenario reserve and the 2% of the scheduled benefits payable within the next 12 months from the date of valuation. Eom said the stochastic reserve would then be the CTE 70 of the floored scenario reserves. Slutsker said the stochastic reserve option could look like calculating a full stochastic reserve for each contract, then doing a CTE 70 of the scenario reserves and CTE 70 for each contract independently, then applying the 2% flooring. Bayerle said he had two concerns: 1) the floor for each scenario will be different, so there will be no comparison across the scenarios, and 2) the open decision on VM-22 regarding the CSV flooring. Bayerle said the Subgroup should be intentional and deliberate about the flooring mechanisms and be considerate of the layering of the different types of floors under VM-22.

Bayerle said ACLI recommended doing all comparisons in aggregate by calculating the CTE 70 in full, then comparing it to the 2% of benefits across all contracts. Bayerle said the ACLI recommended applying the 2% flooring approach at an aggregate level rather than the treaty level because it would make the reconciliation easier as well as align with how other parts of the requirements work. Bayerle said the ACLI could work with the Subgroup to come up with examples of how the different flooring options would work. Slutsker said the discussion would continue on a subsequent call after reviewing examples.

Having no further business, the VM-22 (A) Subgroup adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/A CMTE/LATF/2025-1-Spring/VM-22 Calls/12 11/Dec 11 Minutes.docx

Draft: 2/23/25

Valuation Manual (VM)-22 (A) Subgroup Virtual Meeting December 4, 2024

The VM-22 (A) Subgroup of the Life Actuarial (A) Task Force met Dec. 4, 2024. The following Subgroup members participated: Ben Slutsker, Chair (MN); Elaine Lam, Vice Chair (CA); Lei Rao-Knight (CT); Mike Yanacheak (IA); Vincent Tsang and Matt Cheung (IL); William Leung (MO); Seong-min Eom (NJ); Bill Carmello (NY); Rachel Hemphill and Iris Huang (TX); Tomasz Serbinowski (UT); and Craig Chupp (VA).

1. Discussed Comments on Policyholder Behavior Assumption for Partial Withdrawals Under SPA

Lam said several of the comments received were related to defining the partial withdrawal assumptions for contracts in the accumulation reserving category that have exercised the living benefit guarantees. Lam said the drafting group reexamined the data and recommended deleting the standard projection amount (SPA) assumptions for partial withdrawals related to contracts with guaranteed living benefits (GLBs) after the guarantee has been exercised. She said the intention was for contracts with GLBs to be based on the guaranteed maximum annual withdrawal amounts that are defined within the guarantees' parameters.

Brian Bayerle (American Council of Life Insurers—ACLI) presented comments regarding Section 6.C.4.a through Section 6.C.4.c that asked the Subgroup to clarify the intent. Lam said that in response to the comments, edits were made to clarify the intent and applicability. She said Section 6.C.4.a applies to contracts without a GLB, or if there is a GLB, before it is exercised. She said Section 6.C.4.b edits were made to standardize terminology that clarified that the partial withdrawal amount for contracts with lifetime GLBs is the full guaranteed maximum annual withdrawal amount when the account value is 0. She said Section 6.C.4.c was meant to address all contracts with GLBs. She said 6.C.4.c defined when the contract would commence withdrawals under the GLB.

Lam said commenters raised concerns with the method in Section 6.C.4.c, which directed companies to compare their prudent best estimate utilization assumptions to the prescribed table. She said the draft was updated to reflect the drafting group recommendation to eliminate the comparison and, for simplicity, to use the prescribed table that represented a cumulative utilization rate based on qualification status and age.

Lam noted clarifications in Section 6.C.4.d addressed a subset of policies that have not "exercised" the GLB but took a withdrawal in the contract year immediately preceding the valuation date. She said the requirements assume going forward the benefits received would be the maximum partial withdrawal amount.

Lam said there were some concerns about the qualified and non-qualified utilization assumptions. Lam stated that the data underlying the assumptions showed qualified contracts exhibit higher utilization upon the older ages where the retirement minimum distribution (RMD) ages are involved.

2. Exposed the Updated VM-22 SPA Draft

Lam said edits were made to clarify which assumption to use when the account value is zero. The intent was to create a lapse assumption of zero, and therefore, the minimum and maximum lapse assumptions would not apply when the account value is zero.

Lam noted the ACLI recommended using cash surrender value instead of the account value for the in-themoneyness (ITM) factor for the dynamic lapse formula. Lam said the use of account value was intentionally consistent with VM-21, and she is not recommending a change that would introduce inconsistency with VM-21.

Lam discussed an ACLI comment that the guaranteed minimum interest rate (GMIR) factor in the dynamic lapse formula should be based on the maximum of the guaranteed crediting rate and the underlying GMIR due to the material difference for fixed deferred annuities during their surrender charge period. Lam said the intention was for companies to use the GMIR and not the guaranteed rate because the GMIR could play a big role in contracts with longer expected durations.

Lam said references to fixed indexed and fixed annuities were simplified to indexed annuities and fixed annuities, respectively.

Lam discussed an additional clarification made for indexed annuities regarding the crediting rate definition. She said the edit to use "the options budget or the value of the options supporting the index crediting strategy" was made to address when companies have guaranteed caps.

Lam said the lapse skew application should be consistent with the company's best estimate since it is not an assumption that the policyholder behavior assumptions drafting group looked at.

Bayerle said the ACLI would look at the deferred income annuity and fixed indexed annuity assumptions out of the dynamic lapse formula and compare them with the Milliman data. Bayerle said the ACLI believed the Milliman data did not support the "cliff approach" lapse rates out of the dynamic lapse formula. Lam noted the drafting group had not seen the Milliman data. Lam said she did not recommend changes to the formula; however, the group was open to further discussion.

Andy King (Oliver Wyman) asked why the guaranteed minimum withdrawal benefit (GMWB) utilization assumption used attained age instead of the policy year considering they typically see companies use policy year instead of attained age for these types of guarantees. King said companies could have very high utilization for the younger policies if they had a lot of older contracts. Lam noted that the attained age structure was due to simplification.

Cheung asked for clarification regarding how companies should apply Sections 6.C.4.c and 6.C.4.d for a block of contracts where a portion of people had an immediate withdrawal and that proportion was more than the utilization rates in Table 6.4. He said it made sense to model those that immediately withdraw to continue to withdraw. Chueng said it was unclear how to handle the remaining contracts which have not started yet. He asked if the utilization rate should be zero for those because the utilization rate was already exceeded, or if Table 6.4 applied. He suggested the Subgroup revisit the data and see how a utilization assumption would apply to the portion of the contracts that had not commenced withdrawals. Lam said she would take that question back.

Lam made a motion, seconded by Yanacheak, to expose the updated VM-22 SPA draft, which included edits for partial withdrawal and full surrender SPA assumptions, for a 60-day public comment period end February 7th, 2025. The motion passed.

Having no further business, the VM-22 (A) Subgroup adjourned.

SharePoint/NAIC Support Staff Hub/Member Meetings/A CMTE/LATF/2025-1-Spring/VM-22 Calls/12 04/Dec 4 Minutes.docx

VM-22 Field Test: Preliminary Summary of Participant Results

March 22, 2025 Life Actuarial (A) Task Force



Academy and EY Collaboration: Aggregating the Field Test Results

- The VM-22 field test results have been independently aggregated, clarified, and aligned by the Academy and EY.
- EY contacted submitters, gaining valuable insight.
- Today's results, as presented by EY, reflect the collaborative effort and EY's leadership in the final stages of analysis.
- This presentation represents the publicly discussable results.
- Regulator-only briefings can be scheduled, should that be desired.



Disclaimers

- All participant data received is treated confidentially.
- Participating companies noted varying levels of simplification used to produce field test results within the submission timeframe. Examples include using placeholder assumptions/margins, simplified asset portfolios, only running the Stochastic Reserve and not the Standard Projection Amount, and aggregating inconsistently with proposed VM-22 requirements. Best efforts have been made to analyze and aggregate data submitted by participants. The accuracy and reliability of the results are ultimately dependent on the quality of participant submissions.
- To maintain anonymity of participants per Academy standards, data and metrics for categories with fewer than five participants will not be shared publicly.



Overview and Status

Field Test Participant Results

The purpose of this presentation is to provide a preliminary summary of the VM-22 field test participant results.

This first presentation of results focuses on reserves, including overall impacts, sensitivities, and SERT results.

Where applicable, model office results are shown for comparison or to supplement the field test participant results.

VM-22 field test key objectives



Measure the impact on actual business of the proposed reserve and capital frameworks relative to the current standards to ensure frameworks are working as intended.

Ensure pillars of framework are met

- Appropriate Reflection of Risk
- Comprehensive
- Consistency Across Products
- Practicality and Appropriateness

Test the impact of key open VM-22 design decisions



- AggregationReinvestment guardrail mix
- Stochastic Exclusion Ratio Test threshold
- Standard Projection Amount (SPA) assumptions



Timeline





Field Test Results



VM-22 Participant Data Submitted

The tables below show the counts of companies which submitted results for different components of the field test. Note that cells shaded in gray represent data sections which failed to reach the five-count threshold, resulting in limitations to the analysis presented in the following slides to uphold participant anonymity.

Product	Overall results	SPA results	Margin sensitivities	Reinvestment sensitivities
SPIA	8	5	2	3
PRT	6	4	1	2
SSC	5	4	1	2
FDA (no WB)	11	6	6	6
FDA (WB)	4	3	1	2
FIA (no WB)	12	7	6	6
FIA (WB)	12	6	5	5



Limitations in Participant Results

The accuracy and reliability of the field test results are ultimately dependent on the quality of participant submissions. There were a wide variety of limitations noted from participants which could result in materially different impacts of VM-22 once fully implemented. Below are some of the common limitations that were observed.

1. Assets

The Field Test is showing that assets are one of the key drivers of VM-22 results. Many participants used a simplified approach to allocate assets for the field test, which could have a significant impact on results in some cases. Before applying VM-22 in the future, we expect that companies will perform more analysis and refine their approach to determine the assets that will be used to back VM-22 business, potentially aligning both the asset types and duration matching to the prospective VM-22 business.

2. Standard Projection Amount

Some companies did not provide SPA results or provided SPA results on a different level of aggregation than the SR and therefore could not be analyzed on a product level. Because of this, the overall VM-22 impact from CARVM could be misestimated for those companies.

For the companies that did provide SPA results, there were some inconsistencies in the application of the prescribed assumptions. These were discussed throughout the field test Q&A process and have since been clarified in the requirements.

3. Assumptions and Margins

Many companies noted using placeholder assumptions and/or margins for the field test, and that they plan to do additional analysis to set PBR prudent estimate assumptions for VM-22.

4. PIMR

There was inconsistent treatment of PIMR across participant results. Some companies explicitly disclosed PIMR, some included it in the final reserve, some did not reflect PIMR at all. The summary of results is based on the final VM-22 reserve that participants provided.

5. Aggregation

There were some inconsistencies in the way companies aggregated results, for example including GLWB payout streams in the payout category rather than the accumulation category.

6. Business Included

The field test specification asked for at least 10 years of inforce. Some companies provided less than 10 years (e.g. if the product hasn't been sold for that long), and some companies provided significantly more than 10 years of inforce.



Introduction to the Overall VM-22 Results Slides

- Splits by product:
 - Payout Category: SPIA, PRT, and SSC
 - Accumulation Category: FDA (no WB), FIA (no WB), FIA (WB)
- Model office results for each product
- Total number of companies providing results
- Change in final VM-22 reserve compared to CARVM
 - Mean
 - Median
 - Standard deviation
 - Range

*Final VM-22 reserve = Stochastic Reserve + ASPA – PIMR (when provided). Some companies did not reflect PIMR in the results provided with the field test.



SPA vs. SR by Product

The tables below shows summary statistics comparing the ratio of SPA (including buffer) to the SR. A positive % indicates that the SPA is greater than the SR, while a negative % indicates that the SPA is less than the SR.

The SR is driving the final reserve more often than the SPA for most products, including those that could not be aggregated.

It is expected that most companies will refine the assumptions and margins used for the field test before adopting VM-22, which could have a significant impact on the results below.

Participant results—SPA vs. SR for VM-22

Product	Model office impact	Total # of companies	# SPA =< SR	# SPA > SR	Mean	Median	Standard deviation	Range
SPIA	-2.3%	5	5	0	-2.4%	-2.7%	1.1%	3.1%
FDA (no WB)	-1.0%	6	5	1	-0.8%	-1.0%	1.6%	5.0%
FIA (no WB)	1.6%	7	5	2	-3.1%	-3.0%	6.0%	18.6%
FIA (WB)	3.4%	6	3	3	1.3%	0.0%	4.0%	12.6%

Product Overview

Observations

- The SPA is expected to highlight outliers, so it is not surprising to see the SR dominate for most products.
- Where SPA dominates, it is challenging to pinpoint what the driver is, and whether that is due to simplifications for the Field Test, or whether that is a legitimate outcome in the results. For the WB block, it is believed that the choice in lapse assumptions drove the results in the Field Test.



American Academy of ACTUARIES

SPA vs. SR by Product

Table 1: Participant results—Unbuffered SPA vs. SR for VM-22

Product	Total # of companies	# SPA =< SR	# SPA > SR	Mean	Median	Standard deviation	Range
SPIA	5	5	0	-2.4%	-2.7%	1.1%	3.1%
FDA (no WB)	6	5	1	-0.8%	-1.0%	1.6%	5.0%
FIA (no WB)	7	5	2	-3.1%	-3.0%	6.0%	18.6%
FIA (WB)	6	3	3	1.3%	0.0%	4.0%	12.6%

Table 2: Participant results – Buffered SPA vs. SR for VM-22

Product	Total # of companies	# SPA =< SR	# SPA > SR	Mean	Median	Standard deviation	Range
SPIA	5	5	0	-2.5%	-2.8%	1.1%	3.0%
FDA (no WB)	6	5	1	-0.9%	-1.1%	1.6%	4.9%
FIA (no WB)	7	5	2	-3.6%	-3.2%	5.9%	18.3%
FIA (WB)	6	4	2	0.5%	-0.8%	3.9%	12.2%

Observations

- The tables here show summary statistics comparing the ratio of standard projection amount with and without buffer to the stochastic reserve.
- A positive % indicates that the SPA is greater than the SR, while a negative % indicates that the SPA is less than the SR.
- The SR is driving the final reserve more often than the SPA for most products, including those that could not be aggregated.
- Applying the buffer impacted the dominant reserve for one company.



Overall VM-22 Results: Payout Category

The tables below shows summary statistics on the change from CARVM to the final VM-22 reserve* for field test participants, as compared to the model office results shared previously. Participant results have been normalized so there is equal weighting across companies.

Product	Model office impact	Total # of companies	Mean	Median	Standard deviation	Range		
SPIA	-3.4%	8	-3.3%	-0.9%	13.8%	44.8%		
PRT	-3.5%	6	-0.4%	-1.0%	4.7%	13.2%		
SSC	-5.7%	5	20.9%	9.7%	30.1%	83.1%		

Participant results—CARVM vs. VM-22

Observations

Product Overview

- Model office results show a decrease in VM-22 reserves compared to CARVM, largely driven by work done in the model office to optimize the assets backing the liabilities.
- Wide range of results seen by participants, with some showing an increase in reserves under VM-22.
- From discussions with participants, this is believed to largely be driven by the selection of assets as multiple companies noted they did not spend significant time selecting or optimizing the asset portfolio for the field test.
- PRT saw a tighter range overall, which is believed to be because PRT assets are usually optimized and allocated to specific PRT deals.
- The model office grouped PRT and SSC together, so they are not directly comparable to the participant results.

*Final VM-22 reserve = Stochastic Reserve + ASPA – PIMR (when provided). Some companies did not reflect PIMR in the results provided with the field test.



Overall VM-22 Results: Accumulation Category

The tables below shows summary statistics on the change from CARVM to the final VM-22 reserve* for field test participants, as compared to the model office results shared previously. Participant results have been normalized so there is equal weighting across companies.

FIGUACE OVERVIEW								
Product	Model office impact		Total # of companies	Mean	Median	Standard deviation	Range	
FDA (no WB)	0.3%		11	2.6%	1.6%	4.6%	17.7%	
FIA (no WB)	4.6%		12	6.3%	3.9%	7.9%	27.9%	
FIA (WB)	-16.7%		12	-4.5%	-5.0%	8.4%	26.5%	

Participant results—CARVM vs. VM-22

Observations

Droduct Overviou

- For FDA and FIA (no WB), most companies saw a modest increase while some saw modest decreases. From discussions with individual companies, the main driver appears to be how much effort participants put into asset optimization for the field test.
- As noted previously, the model office for FIA includes a modeling limitation related to the hedge costs and payoffs.
- Most companies saw a decrease compared to CARVM for FIA (WB). This was expected given the treatment for WB riders under CARVM.
- Some companies with FIA (WB) saw an increase, or more modest decrease. From some discussions with participants this may be
 explained by modeling simplifications and/or asset optimization.

*Final VM-22 reserve = Stochastic Reserve + ASPA – PIMR (when provided). Some companies did not reflect PIMR in the results provided with the field test.



Overall VM-22 Results: All Products

The tables below shows summary statistics on the change from CARVM to the final VM-22 reserve* for field test participants, as compared to the model office results shared previously. Participant results have been normalized so there is equal weighting across companies.

Product Overview

Product	Model office impact
SPIA	-3.4%
PRT	-3.5%
SSC	-5.7%
FDA (no WB)	0.3%
FIA (no WB)	4.6%
FIA (WB)	-16.7%

Participant results—CARVM vs. VM-22

Total # of companies	Mean	Median	Standard deviation	Range
8	-3.3%	-0.9%	13.8%	44.8%
6	-0.4%	-1.0%	4.7%	13.2%
5	20.9%	9.7%	30.1%	83.1%
11	2.6%	1.6%	4.6%	17.7%
12	6.3%	3.9%	7.9%	27.9%
12	-4.5%	-5.0%	8.4%	26.5%

*Final VM-22 reserve = Stochastic Reserve + ASPA – PIMR (when provided). Some companies did not reflect PIMR in the results provided with the field test.



High-level Observations Summary

Range of results: There was a wider range of results than was expected across all products. Every
product had at least one company with an increase and one company with a decrease in reserves.

Selection of assets: The assets used in VM-22 modeling are a key driver of results for all products. Given the simplified approaches that many companies took for assets, results could change materially when asset portfolios are refined. Some participants noted that the reinvestment guardrail had a significant impact on results vs. modeling their company reinvestment strategy.

Dominant reserve: Where SPA results were provided, the SR is winning more often than the SPA for payouts and non-WB accumulation products. The SPA is winning more often on WB products. This is likely due to the SPA lapse assumption for WB products. 8 of the total 19 entities that participated in the field test did not provide SPA results.

Notable differences from model office results:

SSC—The model office included SSC as a subset of the PRT block but did not consider SSC as a standalone product so it's not directly comparable to participant results. SSC results also vary depending on the mix of business and inforce duration of the block, which for some participants was much longer than 10 years.

FIA—The model office results included a topside adjustment for the cost of FIA hedges due to a limitation in GGY Axis.



Sensitivity Results Summary



Introduction to the Sensitivities

- The Field Test Specs asked participants to set, and disclose with results, each sensitivity's impact from mortality, policyholder behavior, expenses, hedging, non-guaranteed elements (NGEs), withdrawals, and other assumptions as deemed necessary.
- Participants were also allowed to use some default margins as described in the Specs if they did not want to use their own margins.
- There was only enough information gathered for mortality, lapse rates, expenses, and the reinvestment guardrail; these are discussed on the following slides.
- Similar to the overall results, there are a number of limitations related to sensitivities, e.g., how companies stepped into and isolated each sensitivities impact.


Margin Sensitivities—Mortality

Background

- Field test participants were asked to remove each liability margin individually and provide sensitivity test results.
- The field test specifications included default margins that companies could choose to use in place of their own margins. For mortality, the default margin was +/- 10%.
- Four out of the seven companies included in the analysis below used the default margin. For those who used their own company margins, the margins were <10%.

Results and observations

- Many participants did not provide sensitivity results due to lack of time and resources for the field test.
- Mortality margins were more impactful on accumulation products with WB vs. those without WB, but generally not material for accumulation products overall. Results for the payout category could not be shared publicly, but for the companies that provided results they were largely in line with the WB product results.

Product	# of companies	# of products	# of products > 0%	# of products =< 0%	Mean	Median	Standard deviation	Range
FA (no WB)	5	5	0	5	-0.03%	-0.01%	0.04%	0.11%
FIA (no WB)	5	5	0	5	-0.15%	-0.02%	0.21%	0.55%
FIA (WB)	5	5	1	4	-1.01%	-1.13%	0.92%	2.41%
FA + FIA (WB)	6	6	1	5	-0.97%	-0.96%	0.85%	2.41%



Margin Sensitivities—Lapse

Background

- Field test participants were asked to remove each liability margin individually and provide sensitivity test results.
- The field test specifications included default margins that companies could choose to use in place of their own margins. The margins provided were +/- 10% on base lapse and +/- 150% on dynamic lapse.
- Three out of the seven companies included in the analysis below used the default margin. For those who used their own company margins, one out of the seven used margins >10% and three out of the seven used margins <10%.

Results and observations

- Many participants did not provide sensitivity results due to lack of time and resources for the field test. For those that did provide results, we were able to aggregate the results of a base lapse sensitivity as shown below. Very few companies provided sensitivity testing on dynamic lapses and therefore results could not be aggregated.
- The base lapse margin sensitivity had an immaterial impact for most companies.

Product	# of companies	# of products	# of products > 0%	# of products =< 0%	Mean	Median	Standard deviation	Range
FA (no WB)	6	6	0	6	-0.43%	-0.27%	0.49%	1.48%
FIA (no WB)	6	6	1	5	-0.62%	-0.03%	1.32%	3.57%
FIA (WB)	5	5	1	4	-0.64%	-0.05%	1.10%	2.85%
FA + FIA (WB)	6	6	1	5	-0.54%	-0.05%	1.03%	2.85%

Margin Sensitivities—Expenses

Background

- Field test participants were asked to remove each liability margin individually and provide sensitivity test results.
- The field test specifications included default margins that companies could choose to use in place of their own margins. For lapse, the default margin was +/- 5%.
- Three out of the five companies included in the analysis below used the default margin. For those who disclosed their own company margins, the margins were <5%.

Results and observations

- Many participants did not provide sensitivity results due to lack of time and resources for the field test. The results below are aggregated across FA and FIA products without WB. We received limited results for other products that could not be aggregated, however the results were consistent across all products.
- The expense margin sensitivity had an immaterial impact for all participating companies.

Product	# of companies	# of products	# of products > 0%	# of products =< 0%	Mean	Median	Standard deviation	Range
FA + FIA (no WB)	5	8	1	7	-0.01%	-0.02%	0.01%	0.04%



Reinvestment Guardrail Sensitivity

Background

Field test participants were asked to provide results for two reinvestment guardrail sensitivities:

- Baseline: 50% AA, 50% A
- Required Sensitivity: 5% Treasury, 15% AA, 40% A, 40% BBB
- Optional Sensitivity: 5% Treasury, 15% AA, 80% A

Results and observations

- Many participants did not provide sensitivity results due to lack of time and resources for the field test.
- Overall, the reinvestment guardrail sensitivities did not have a material impact on reserves for most companies. Five of the seven companies included in the below analysis had an impact of <1% for all products.
- The results below show the impact of the required sensitivity vs. baseline for products where we had a sufficient number of data points to aggregate results:

Product	# of companies	# of products	# of products > 0%	# of products =< 0%	Mean	Median	Standard deviation	Range
FA (no WB)	6	6	2	4	-0.13%	-0.05%	0.18%	0.51%
FIA (no WB)	6	6	1	5	-0.10%	0.00%	0.30%	0.96%
FIA (WB)	5	5	1	4	-0.41%	-0.54%	0.46%	1.17%
FA + FIA (WB)	6	7	2	5	-0.29%	-0.34%	0.46%	1.21%



Stochastic Exclusion Ratio Test



Stochastic Exclusion Ratio Test

Background

• Field test participants were asked to perform the Stochastic Exclusion Ratio Test (SERT) as outlined in the proposed VM-22 requirements.

Results and observations

- Many participants chose not to provide SERT results due to several factors:
 - Lack of resources to produce results in time for the field test
 - Working assumption that their business would not pass the SERT and therefore they do not plan to run it
 - Do not plan to run the SERT because they want to calculate VM-22 stochastic reserves
- Several companies provided partial results but not enough information to calculate the final SERT ratio. <u>If any participating</u> companies have this information available but did not submit it already, please reach out.
- As a result, field test participant SERT results could not be aggregated and shared publicly.
- For the limited data points provided, the participant SERT results were consistent with the model office results.
- Out of the 11 companies that submitted at least partial results, 10 of them used a mortality margin of +/- 5%, while 1 of them opted to use a mortality margin of +/- 10%.
- The model office SERT results (presented previously) are included on the following slide for reference.



Stochastic Exclusion Ratio Test – Model Office

The table below summarizes the model office results of the stochastic exclusion ratio test for each product. The results in each column show the resulting ratio when "b" from the SERT ratio calculation* is calculated under the given mortality sensitivity.

The impact of applying a +/- 5% mortality margin did not materially impact the resulting ratio for the accumulation products.

Product	95% mortality factor	100% mortality factor	105% mortality factor	
SPIA	5.6%	3.3%	1.2%	
PRT	6.0%	3.4%	1.0%	
FDA (no WB)	1.4%	1.3%	1.2%	
FDA (WB)	2.2%	2.2%	2.1%	
FIA (no WB)*	5.8%	5.8%	5.8%	
FIA (WB)*	33.8%	33.7%	33.6%	

*Exclusion test ratio = (b - a) / c

- a = adjusted reserve under the baseline scenario
- b = largest adjusted reserve under the 16 prescribed scenarios
- c = present value of benefits under the baseline scenario

*Important disclaimer for the FIA model office results: the cost of the FIA hedges is currently accounted for via a spreadsheet topside for each scenario. The model currently incorporates the payoffs of the hedges, but not the costs. We have included the costs via topside, estimated as option budget x AV / 12 (since there are annual resets), which are reflected in the results above and throughout this presentation. A system enhancement is in progress from the vendor.

American Academy of Actuaries

Capital Results Summary



Participant Capital Results: Change in C-3 RBC

The tables below shows summary statistics on the percentage change from the old C-3 calculation to the proposed C-3 approach included in the field test instructions. Participant results have been normalized so there is equal weighting across companies. All participants used the MTA approach:

YY% x ((CTE (XX) + [Additional Standard Projection Amount] – Statutory Reserve) x (1 – Federal Income Tax Rate) – (Statutory Reserve – Tax Reserve) x Federal Income Tax Rate)

Participant results—Old C-3 vs. New C-3

Statistic	Products	# of companies	XX = 98% YY = 30%	XX = 98% YY = 25%	XX = 98% YY = 20%	XX = 95% YY = 30%	XX = 95% YY = 25%	XX = 95% YY = 20%
Mean	All	13	52%	26%	1%	-16%	-30%	-44%
Median	All	13	-5%	-21%	-37%	-59%	-66%	-72%
Standard Deviation	All	13	153%	128%	102%	104%	87%	69%
Range	All	13	501%	418%	334%	387%	322%	258%

Observations

- Companies provided capital results with varying levels of aggregation, which made it difficult to summarize results in a way that could be shared publicly. The results above summarize the total capital impact for each company, which in some cases includes a single product and others include five+ products. Some companies reflected aggregation benefits in their capital calculations while others did not.
- C3P1 results are based on AIRG scenarios while the proposed capital results are based on the same proposed GOES scenarios that were used for the VM-22 calculations in the field test.
- Companies with only accumulation products tended to see more decreases in capital, however there were a wide range of results for all product combinations.
- The results are heavily skewed by a few companies with large increases in capital. On the following slide, the summary of results is broken down for companies that had an increase vs. companies that had a decrease in reserves under the proposed XX=98% and YY%=25%.



Participant Capital Results: Change in C-3 RBC

Participant results—Old C-3 vs. New C-3 for Companies with a Decrease in Capital under XX=98%, YY=30%

Statistic	Products	# of companies	XX = 98% YY = 30%	XX = 98% YY = 25%	XX = 98% YY = 20%	XX = 95% YY = 30%	XX = 95% YY = 25%	XX = 95% YY = 20%
Mean	All	7	-62%	-68%	-75%	-92%	-94%	-95%
Median	All	7	-77%	-80%	-84%	-96%	-97%	-98%
Standard Deviation	All	7	33%	28%	22%	19%	16%	13%
Range	All	7	99%	82%	66%	68%	57%	45%

Participant results—Old C-3 vs. New C-3 for Companies with an Increase in Capital under XX=98%, YY=30%

Statistic	Products	# of companies	XX = 98% YY = 30%	XX = 98% YY = 25%	XX = 98% YY = 20%	XX = 95% YY = 30%	XX = 95% YY = 25%	XX = 95% YY = 20%
Mean	All	6	184%	137%	89%	72%	44%	15%
Median	All	6	173%	128%	82%	24%	3%	-17%
Standard Deviation	All	6	130%	108%	87%	92%	77%	61%
Range	All	6	362%	302%	242%	260%	217%	173%



Participant Capital Results: Comparison of CTE levels

The overall impacts from the current C-3 calculation to the proposed C-3 approach varied widely, largely due to the variances in treatment under current capital. On this slide, the table below shows summary statistics comparing CTE(98) and CTE(95) against CTE(70) to demonstrate the distribution of participants' results in the tails. Participant results have been normalized so there is equal weighting across companies.

Participant Results—CTE(XX)

Measure	Products	Number of companies	Mean	Median	Standard Deviation	Range
Percent change from CTE(70) to CTE(98)	All	13	4.3%	4.3%	2.6%	8.0%
Percent change from CTE (70) to CTE(95)	All	13	2.9%	1.9%	1.8%	5.4%

Observations

- Companies with larger tail risk—e.g. higher CTE(98) relative to CTE(70)—tended to have increases in capital under the proposed method as compared to old C-3.
- CTE(95) results were right-skewed, meaning there were some companies with large increases in relation to CTE(70) which increased the mean relative to the median. CTE(98) was more evenly distributed, but with a wider range of results. There is more variability in CTE(98) vs CTE(95), which is expected given the more extreme tail risk being considered.
- Companies with products from the payout category tended to see higher tail risk, however there was a range of results across all products.



Questions or Comments:

Amanda Barry-Moilanen Policy Project Manager, Life <u>barrymoilanen@actuary.org</u>

Steve Jackson Director of Research <u>sjackson@actuary.org</u>



Agenda Item 4

GOES Presentation

2025 SPRING NATIONAL MEETING/INDIANAPOLIS

Agenda

- 1. APF 2025-04: Specific Topics
 - a) Scenario Subset Requirements
 - b) Stochastic Exclusion Ratio Test (SERT) Threshold
 - c) Stochastic Exclusion Test (SET) Volatility
 - d) SERT Basis
- 2. ACLI Proposed DR Scenario Revisions
 - a) Appendix: Evaluating Approach for Different Tenors - Current DR Approach
- 3. GOES Model Governance Framework Key Topics
- 4. Next Steps
- 5. Appendix Additional Field and Model Office Testing Results

50





2025 SPRING NATIONAL MEETING / INDIANAPOLIS



APF 2025-04: Scenario Subset Updates

- Conning has delivered an excel-based tool that selects subsets from the 10,000-scenario file based on UST significance values or Large Cap equity fund gross wealth factors (GWFs).
- VM-20 currently prescribes the use of the scenario picker built into the AIRG, which uses UST significance values to produce 1,000; 500; 200; or 50 scenario subsets.
- VM-21 allows for fewer than 1,000 scenarios to be used provided they materially reproduce the CTE results from running a larger scenario set.
- The new VM-20 and VM-21 language would allow for the use of scenario subsets provided they meet the simplification, approximation, and modeling efficiency technique requirements of VM-20 section 2.G and VM-21 3.H
- Question: Should scenario selection be moved under the simplifications, approximations, and modeling efficiency techniques requirements?

VM-20, Section 7.G.2.c

c. Use of fewer scenarios rather than a higher number of scenarios is permissible as a model efficiency technique provided that t:

 The smaller set of scenarios is generated using the scenario picker tool provided within the prescribed scenario generator, and
 the use of the technique is consistent with Section 2.G.

VM-21, Section 8.F

- F. Number of Scenarios and Efficiency in Estimation Use of fewer scenarios rather than a higher number of scenarios is permissible as a model efficiency technique provided that the use of the technique is consistent with Section 3.H.
 - 1. For straight Monte Carlo simulation (with equally probable "paths" of fund returns), the number of scenarios should typically equal or exceed 1000. The appropriate number will depend on how the scenarios will be used and the materiality of the results. The company should use a number of scenarios that will provide an acceptable level of precision.
 - Fewer than 1,000 scenarios may be used provided that the company has determined through prior testing (perhaps on a subset of the portfolio) that the CTE values so obtained materially reproduce the results from running a larger scenario set.
 - 3. Variance reduction and other sampling techniques are intended to improve the accuracy of an estimate more efficiently than simply increasing the number of simulations. Such methods can be used provided the company can demonstrate that they do not lead to a material understatement of results. Many of the techniques are specifically designed for estimating means, not tail measures, and could in fact reduce accuracy (and efficiency) relative to straight Monte Carlo simulation.

Guidance Note: With careful implementation, many variance reduction techniques can work well for CTE estimators. For example, see Manistre, B.J., and Hancock, G. (2003), "Variance of the CTE Estimator," 2003 Stochastic Modeling Symposium, Toronto, September 2003.

4. The above requirements and warnings are not meant to preclude or discourage the use of valid and appropriate sampling methods, such as Quasi Random Monte Carlo (QRMC), importance sampling or other techniques designed to improve the efficiency of the simulations (relative to pseudo random Monte Carlo methods).





Model Office Testing: SERT Analysis

Impacts of scenario revisions and prudent best estimate assumptions

March 12, 2025

A business of Marsh McLennan

CONFIDENTIALITY

Our clients' industries are extremely competitive, and the maintenance of confidentiality with respect to our clients' plans and data is critical. Oliver Wyman rigorously applies internal confidentiality practices to protect the confidentiality of all client information.

Similarly, our industry is very competitive. We view our approaches and insights as proprietary and therefore look to our clients to protect our interests in our proposals, presentations, methodologies, and analytical techniques. Under no circumstances should this material be shared with any third party without the prior written consent of Oliver Wyman.

VM-20 ULSG PROJECTIONS AND MODEL OFFICE DESCRIPTION

Model assumptions and product features were selected based on industry benchmarks to be a simplified representation of products currently offered

Component	Description of functionality
	 Universal life with shadow design lifetime secondary guarantee issued in 2020
	 Time 0 reserves are held in 50% 5-year BBB bonds and 50% 7-year BBB bonds
	 Reinvestment strategy uses 50% A/AA corporate bonds
Projection model details	– 10% 5-year
	– 25% 7-year
	– 35% 10-year
	– 25% 20-year
	– 5% 30-year
	 Follows industry benchmark assumptions
Best estimate assumptions	 Mortality experience is 100% credible with 25 years of sufficient data
	 UL crediting rate is dynamic and based on NAER less a spread, varying for each stochastic scenario
Prudent estimate	 VM-20 prescribed mortality margins based on credibility and sufficient data period
assumptions	 Minimal lapse when policy maintained inforce by NLG (i.e. CSV = 0)

6

REVISIONS TO GOES

Initial Treasury Yield Curve Fitting Methodology: The revised initial yield curve fitting
 methodology places more emphasis on the longer maturities for greater alignment with insurance company investment strategies.

Dynamic Generalized Fractional Flooring (DGFF): The DGFF methodology is an extension of
 the previous generalized fractional floor and the parameters are set to target a 3% level of
 negative 1-year UST rates in the steady state.

Equity Calibration: The revised equity calibration raises the 1st percentile gross wealth factors
 (GWFs) of the Large Capitalization equity fund to be closer to the acceptance criteria targets compared to the prior 2024 field test calibration.

Note: Z3 scenario changes do not affect the VM-20 model office results since the product is not linked to equity markets

SERT RESULTS – BASELINE

SERT results were tested using AIRG and GOES scenario sets. An additional test was performed using prudent best estimate assumptions

SERT ratios and underlying components



Scenario Set	Assumption	Baseline adj DR	Max adj DR (#3 pop down)	Total PV benefits	
		[a]	[b]	[c]	
AIRG	AE	1,013,170	1,555,310	6,296,790	
Z0	AE	1,077,760	2,264,180	6,523,710	
Z3	AE	999,528	2,207,050	6,375,300	
Z3	PBE	1,690,910	3,578,650	7,543,120	

Commentary

- Transitioning from AIRG to Z0 scenarios significantly increases the SERT ratio due to a large rise in the maximum adjusted DR, outweighing a slight increase in PV benefits.
- Conning scenario revisions between Z0 and Z3 led to a minor increase to the SERT ratio, driven by lower PV benefits and a widened spread between baseline DR and max adjusted DR
- Changing assumptions from AE to PBE leads to a significant increase to the SERT ratio, driven by a significantly widened spread between baseline DR and max adjusted DR outweighing higher PV benefits
- Z0 = 2024 GOES Field Test Scenario Set 1
- AE = anticipated experience assumptions w/o margins
- PBE = prudent best estimate assumptions w/ margins

7

SERT ratios were most heavily impacted by the initial change from AIRG to Conning scenarios and the application of PBE assumptions

NAIC

2025 SPRING NATIONAL MEETING / INDIANAPOLIS

2024 Field Test Participant SERT Results

- 87% of the field test participants' model segments passed the SERT in their baseline YE23 run with a 6% threshold. This number dropped to 77% for the FT1 YE23 scenarios. Increasing the threshold to 7% brings the participant passing rate back up to a similar level.
- 58% of the FT3 (Up Rate Shock) field test participant model segments passed the SERT at the 6% threshold, increasing to up to ~80% if the threshold is increased to 7.5%. Note, we do not have comparative data on the frequencies of participants that would pass using the FT3 starting yield curve and AIRG SERT scenarios.

Frequency of Passing SERT by Field Test Run All VM-20 Reserving Categories



2025 SPRING NATIONAL MEETING/INDIANAPOLIS

NAIC

APF 2025-04: Stochastic Exclusion Ratio Test

- 2024 field test data showed average SERT results increase for participants, with some additional failures resulting from switching to GOES
- Question: Should the SERT scenario threshold be revised given the increase to average SERT scenario results?



VM-20 Section 6.A.2.c

- c. If the ratio calculated in Section 6.A.2.a above is less than 6[X]% pre-YRT reinsurance, but is greater than 6[X]% post-YRT reinsurance, the group of policies will still pass the SERT if the company can demonstrate that the sensitivity of the adjusted DR to economic scenarios is comparable pre- and post-YRT reinsurance.
 - 1. An example of an acceptable demonstration:
 - For convenience in notation SERT = the ratio (b-a)/c defined in

 (a) above
 - The pre-YRT reinsurance results are "gross of YRT," with a subscript "gy," so denoted SERT_{gy}
 - The post-YRT results are "net of YRT," with subscript "ny," so denoted SERT_{ny}
 - b. If a block of business being tested is subject to one or more YRT reinsurance cessions as well as other forms of reinsurance, such as coinsurance, take "gross of YRT" to mean net of all non-YRT reinsurance but ignoring the YRT contract(s), and "net of YRT" to mean net of *all* reinsurance contracts. That is, treat YRT reinsurance as the last reinsurance in, and compute certain values below with and without that last component.
 - c. So, if SERT_{gy} ≤ 0.060_0[X]0 but SERT_{ny} > 0.0600[X]0, then compute the largest percent increase in reserve (LPIR) = (b-a)/a, both "gross of YRT" and "net of YRT."

 $LPIR_{gy} = (b_{gy} - a_{gy})/a_{gy} LPIR_{ny} = (b_{ny} - a_{ny})/a_{ny}$

Note that the scenario underlying b_{gy} could be different from the scenario underlying $b_{\pi y}.$

If SERT_{gy} × LPIR_{my}/LPIR_{gy} < 0.0000[X]0, then the block of policies passes the SERT.

2025 SPRING NATIONAL MEETING/INDIANAPOLIS

NAIC

APF 2025-04: Stochastic Exclusion Test Updates

SET Certification Method:

VM-20 Section 6.A.1.a.iii

SET Certification Method – For any groups of policies within the scope of VM-20, the qualified actuary may document that a group of policies has passed the exclusion test through an approach other than the SET Certification Method within the past three years and that there have not been material changes in the interest rate risk or asset return volatility risk inherent in the liabilities and supporting assets. Alternatively, fFor groups of policies other than variable life or ULSG, in the first year and at least every third calendar year thereafter, the company provides a certification by a qualified actuary that the group of policies is not subject to material interest rate risk or asset return volatility risk (i.e., the risk on non-fixed-income investments having substantial volatility of returns, such as common stocks and real estate investments). The company shall provide the certification and documentation supporting the certification to the commissioner upon request.

- SERT results provided by the 2024 GOES
 Field Test participants showed increased
 SERT ratios in the field test runs compared to
 the AIRG baseline. Some participants went
 from passing in their baseline SERT results to
 failing in the field test scenarios.
- Question: Should additional flexibility to the SET be added to address volatility?

Basis:

VM-20 Section 6.A.2.b.i.a
a) The DR defined in Section 4.A, but with the following differences:

Using anticipated experience assumptions with no margins.
Using the interest rates and equity return assumptions specific to each scenario.

3)2) Using NAER and discount rates defined in Section 7.H specific to each scenario to discount the cash flows.

- The SERT results are currently determined using anticipated experience with no margins.
- The 2024 GOES Field Test had one participant pass the SERT but calculate an ST that was higher than their DR or NPR.
- Question: Should the SERT use prudent estimate assumptions?



Deterministic Reserve Scenario Analysis

March 22, 2025

Deterministic Reserve Scenario

- The current deterministic reserve scenario (DR) is designed to have a particular relationship to the stochastic distribution.
 - Uniform shocks over the first 20 years to get down to the 84th percentile of 20-year shocks
 - No shocks after year 20 allowing rates to drift back to median based on mean reversion
- Unlike the AIRG, the Conning model does not have a straightforward way to replicate this, and the current approach is producing a DR that falls materially lower in the stochastic distribution of rates.
 - Focus is on relationship to stochastic distribution
 - Conning DR will be lower in absolute rates than AIRG DR. That is expected and not the issue.
- This would throw off the original design of the relationship between the deterministic reserve & stochastic reserve such that the DR would tend to drive final reserves.
 - Overriding the more refined, technically developed stochastic reserve
 - Resulting in higher reserves than industry expectations

Comparison of Percentiles vs Stochastic of AIRG and Updates GOES

Cumulative factor is based on a gross wealth factor type approach CumulativeFactor_t = CumulativeFactor_{t-1} * MonthlyRate^(1/12)



Field Test Deterministic Scenario percentile relationship with the stochastic distribution is materially lower than AIRG.

- The rates in the first 20 years fall below the 16th percentile
- The rates never return to median
- Cumulative rate measure percentiles show compounding effect of consistently falling below the AIRG percentiles, especially in first 20 years

13

Alternative Deterministic Reserve Scenario Approaches

- The goal of changing the DR approach is to re-align DR to their originally designed relationship with stochastic distribution.
- Recommended approach: Adding an adjustment to the underlying shocks to re-align percentiles (formulaic notation to be provided)
 - Current preference would be to add .01 to shocks in all years is current best fit; easier implementation and maintain proper relationships among rates and returns
 - May want to review methodology if updates to calibration of the interest rate model (e.g., updates to the mean reversion parameters) as part of the 5-year review
- Other options explored:
 - 1. Adjusting only the later years to median
 - Leaves early year issues
 - 2. Calculate DR outside of GOES based on percentiles of stochastic scenarios
 - Additional work outside of model; complicates projecting reserves
 - Concern about maintaining proper relationships amongst rates/returns
 - 3. Euler Alternative suggested mathematical method
 - Not recommended: larger deviation vs stochastic

AIRG (Target) vs. Shocks + .01



Adjusting the shocks upwards by defined amounts improves the relationship of the DR to the stochastic distribution. Initial analysis indicates .01 resulting in a closer fit to AIRG percentiles

15

Evaluating Approach for Different Tenors – Shocks + .01



Approach of adding shocks by a flat .01 results in stable percentile relationships

*Comparable graph based on Current DR approach in Appendix

Evaluating Approach Across Scenario Sets – Shocks + .01





Appendix

Evaluating Approach for Different Tenors – Current DR Approach



Current approach consistently falls below targeted AIRG percentiles

19

2025 SPRING NATIONAL MEETING / INDIANAPOLIS



GOES Model Governance: Fallback Plan

Interested Party Comments

- There should be a well-defined communication, escalation, and fallback plan if something unexpected occurs during scenario generation (as well as a business continuation plan for other potential disruptions). Year-end and quarter-end valuations are typically the most critical and require tighter recovery / resolution times.
- In what situations would NAIC pause the release of scenarios? How would the process be handled after escalation to reach resolution?
- ...up to a one-day delay would be acceptable. If there is a significant issue identified after scenarios are posted, then there is a larger discussion to be had beyond having a contingency plan such as allowing companies to revert back to the previous month's scenarios with any appropriate adjustments...

- Different approaches for different situations:
 - Interim month-end vs. quarter/year-end
 - Minor issue in which corrected scenarios can be released on EOM + 2 vs. prolonged posting delay
 - Issue caught during standard Conning/NAIC review or days later by interested party
- Potential Resolutions:
 - Use of prior month-end scenarios
 - Additional day delay to post scenarios with communication to industry
 - NAIC using Conning software to produce and post scenarios

2025 SPRING NATIONAL MEETING/INDIANAPOLIS

NAIC

GOES Model Governance: Scenario Review

Interested Party Comments

- The latest set of acceptance criteria includes "targeting criteria" and "evaluation statistics." These should be defined in the governance framework to clarify how they will be used in determining whether scenario sets are acceptable.
- Validation reports need to show how well the model performs against acceptance criteria and stylized facts. It may be helpful to provide more detail on the process / thresholds to determine whether a scenario set is acceptable, while still allowing for appropriate use of judgment.

- Idea of "Dashboard" has been proposed to allow for the quick review of the scenarios against the acceptance criteria.
- Many of the regulator adopted targeting criteria and evaluation statistics are not dependent on the starting economic environment.
 - For example, Targeting Criteria T1.T defines how many high rates that are permissible. In starting environments with higher interest rates, these criteria may not be appropriate.
- Development of more robust thresholds and dynamic criteria could be considered as a "Day II" item.

2025 SPRING NATIONAL MEETING / INDIANAPOLIS

NAIC

GOES Model Governance: Periodic Updates

Interested Party Comments

- ACLI would suggest regularly scheduled meetings for discussion about whether there is a need for model or calibration updates. Such discussions could also make it easier to identify items for the 5-year recalibration and model revisions (or sooner if deemed necessary) and off-cycle model updates as described in the draft framework.
- The 5-year review should be a comprehensive model review and include assessing the continued suitability of the model form/structure (which may include evaluation of vendor limitations) and not limited to the recalibration of the existing model.
- The 5-year review process should commence well before 5 years has elapsed. (Starting the process in 5 years would delay any update significantly beyond 5 years.)

- A 5-year calibration cycle, annual back testing report, and situational off-cycle updates are all currently envisioned as part of the model governance framework. The annual back testing report could illustrate the need for an off-cycle update.
- Next draft of model governance framework should establish a timeline for work on the five-year recalibration.
- Conning routinely performs research on economic scenario generators and adds or revises features to their economic scenario generator offerings. The timing and process for accepting model enhancements should be clearly defined.

2025 SPRING NATIONAL MEETING/INDIANAPOLIS

GOES Model Governance: Alternative Models

Interested Party Comments

 While not specifically related to governance, we would appreciate a future opportunity to discuss whether proprietary models that comport with the stylized facts and are within the thresholds for targeting criteria and evaluation statistics facts would be acceptable for valuation purposes. We note that developing these thresholds is something ACLI and our team of subject matter experts are currently working so this is a topic we would be willing to present on in the coming weeks.

- Current Valuation Manual requirements allow for the use of alternative economic scenario generators. For example, VM-21 allows for the use of non-prescribed generators provided that the Total Asset Requirement (TAR) is not materially understated.
- Prior to VM-21, proprietary equity models were allowed provided that they met calibration criteria. However, this would be a new concept for the Valuation Manual.
2025 SPRING NATIONAL MEETING/INDIANAPOLIS

NAIC

GOES Model Governance: Proprietary Bond Model

Interested Party Comments

- ACLI members have also identified questions due to the proprietary nature of certain aspects of the model, particularly the Corporate Model.
 - It is common practice for companies to replicate models to help manage their business (e.g., projecting future reserves / capital requirements for capital / risk management). We are concerned that an NDA would restrict this ability; we would be comfortable if the NDA explicitly prohibits the documentation from being used for any purpose external to the company or for commercial purposes.
 - Further, it is unclear how ACLI could have discussions related to aspects of the generator that are proprietary. If not all of our members have signed an NDA, it would seem we would not be able to discuss those items at all. This would also be an issue for any public discussions.

Ideas for Discussion

- Companies are not restricted from developing models that could produce scenarios similar enough to meet their needs.
- Companies could bring issues to Conning, NAIC Staff, and state insurance regulators.
- Conning could consider partial release from the NDA to discuss particular issues on a case-by-case basis.

2025 SPRING NATIONAL MEETING/INDIANAPOLIS

NAIC

Next Steps

Continue Work of the GOES (E/A) Subgroup

- NAIC Staff will utilize feedback to revise the draft Model Governance Framework. A revised draft will be taken to the GOES (E/A) Subgroup to discuss high-priority revisions.
- The GOES (E/A) Subgroup will continue to discuss any remaining implementation items.

Work Towards GOES Adoption

- A joint meeting of the Life Actuarial (A) Task Force and the Life RBC (E) Working Group will be scheduled to confirm key decisions made at the GOES (E/A) Subgroup level.
- The Life Actuarial (A) Task Force will need to adopt APF 2025-04 by mid-year 2025. The Life RBC (E) Working Group will need to adopt blanks changes by mid-year 2026.

Prepare for Implementation

- NAIC Staff and Conning will work to build out governance processes and production scenario posting procedures.
- Documentation will be updated, enhanced, and streamlined into a comprehensive document.



Appendix 2024 GOES Field Test Results

NAIC NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

75

2025 SPRING NATIONAL MEETING / INDIANAPOLIS



2024 Field Test Participant SERT Results

- For the 12/31/23 GOES FT1 scenarios compared to the Baseline (AIRG) SERT scenarios:
 - The average SERT ratio increased across all VM-20 reserving categories, and
 - Each reserving category saw one participant's model segment that had passed with the Baseline fail with the GOES SERT scenarios.
- The average SERT ratio across each reserving category was significantly impacted by increases to the model segment that failed with the Baseline
- FT3 ("Up Rate Shock") saw the most model segments fail, particularly in the term model segment.
- No additional "All Other" model segments failed the field test SERT scenarios

Average Participant SERT Ratio by Reserving Category



Number of Passing Participant Model Segments/Total Participant Model Segments

VM-20 Reserving Category	Baseline	FT1 12/31/23	FT2 Low Rate Shock	FT3 Up Rate Shock	FT4 Normal Yield Curve	FT6 Alt. Initial Yield Curve	e Fit
ULSG	6/7	5/7	6/7	4/7	5/7	4/6	
Term	10/11	8/10	7/8	3/7	5/7	5/7	
All Other	4/5	4/5	4/5	4/5	4/5		
							27

2025 SPRING NATIONAL MEETING / INDIANAPOLIS

NAIC

2024 Field Test Participant SERT Results

- For the Term and ULSG reserving categories, when the model segment that is failing in the baseline is removed:
 - the average SERT ratios go down significantly.
 - the average SERT ratio is never above the passing threshold.
- There were not enough participants to show for the "All Other" VM-20 Reserving Category



APPENDIX – Model Office SERT Scenario Level results

SCENARIO LEVEL RESULTS - ADJUSTED DR

SERT Scenario	AIRG	Z0	Z3	Z3 Prudent
1 – Pop up, high equity	448,508	280,851	223,119	437,636
2 – Pop up, low equity	448,508	280,851	223,119	437,636
3 – Pop down, high equity	1,555,310	2,264,180	2,207,050	3,578,650
4 – Pop down, low equity	1,555,310	2,264,180	2,207,050	3,578,650
5 – Up/down, high equity	830,102	789,954	707,633	1,231,650
6 – Up/down, low equity	830,102	789,954	707,633	1,231,650
7 – Down/up, high equity	1,178,630	1,353,220	1,287,620	2,113,200
8 – Down/up, low equity	1,178,630	1,353,220	1,287,620	2,113,200
9 – Baseline scenario	1,013,170	1,077,760	999,528	1,690,910
10 – Inverted yield curves	930,815	875,015	839,873	1,451,040
11 – Volatile equity returns	1,013,170	1,077,760	999,528	1,690,910
12 – DR scenario	1,300,920	1,689,200	1,604,060	2,618,240
13 – Delayed pop up, high equity	663,608	532,833	474,967	841,649
14 – Delayed pop up, low equity	663,608	532,833	474,967	841,649
15 – Delayed pop down, high equity	1,277,230	1,671,340	1,585,740	2,587,390
16 – Delayed pop down, low equity	1,277,230	1,671,340	1,585,740	2,587,390

© Oliver Wyman

30

SCENARIO LEVEL RESULTS - TOTAL PV BENEFITS

SERT Scenario	AIRG	Z0	Z3	Z3 Prudent
1 – Pop up, high equity	5,247,808	4,954,830	4,808,978	5,374,781
2 – Pop up, low equity	5,247,808	4,954,830	4,808,978	5,374,781
3 – Pop down, high equity	7,236,575	8,468,330	8,379,609	10,362,714
4 – Pop down, low equity	7,236,575	8,468,330	8,379,609	10,362,714
5 – Up/down, high equity	5,891,597	5,866,760	5,699,969	6,641,930
6 – Up/down, low equity	5,891,597	5,866,760	5,699,969	6,641,930
7 – Down/up, high equity	6,652,163	7,118,032	7,002,948	8,356,130
8 – Down/up, low equity	6,652,163	7,118,032	7,002,948	8,356,130
9 – Baseline scenario	6,296,786	6,523,711	6,375,300	7,543,123
10 – Inverted yield curves	6,110,477	6,021,058	5,969,316	7,023,407
11 – Volatile equity returns	6,296,786	6,523,711	6,375,300	7,543,123
12 – DR scenario	6,746,353	7,447,647	7,301,762	8,850,344
13 – Delayed pop up, high equity	5,844,807	5,791,079	5,665,601	6,470,649
14 – Delayed pop up, low equity	5,844,807	5,791,079	5,665,601	6,470,649
15 – Delayed pop down, high equity	6,627,257	7,265,034	7,109,860	8,614,814
16 – Delayed pop down, low equity	6,627,257	7,265,034	7,109,860	8,614,814

© Oliver Wyman

31