

Your Future, Your Super Review Consultation paper

Submission

October 17, 2022

Dr David Bell

About The Conexus Institute

The Conexus Institute is an independent, not-for-profit research institution focused on improving retirement outcomes for Australian consumers. Philanthropically funded, The Institute is supported by the insights of a high-quality advisory board, whereby each member's involvement is on a pro-bono basis. The Institute adopts a research-for-impact model and frequently collaborates with researchers from academia, associations, and industry. Research is generally made open source to create transparency and accountability. The Conexus Institute exists with no commercial relationships. Further information [here](#).

About David Bell

Dr David Bell is Executive Director of The Conexus Institute. Bell's career has been dedicated to the investment and retirement sector. He has worked with both commercial and profit-for-member firms, and ran his own consulting firm. Bell worked with APRA in the development of the APRA Heatmap. Academically, Bell taught for 12 years at Macquarie University and in 2020 completed his PhD at UNSW which focused on retirement investment problems. Full bio [here](#).

***** The author is willing and able to participate in further consultation. *****

1. Summary and recommendations

Most of our work in this submission focuses on the performance test. We consider all aspects of the YFYS reforms, but the performance test is where we believe we can make the greatest contribution.

We believe that consumers, especially those in default options, deserve the protection of a high-quality, effective performance test. Unfortunately, our view is that the current YFYS performance test does not provide appropriate protection to consumers.

The YFYS performance test is backwards-looking and does not account for important qualitative considerations which inform future performance outcomes. The test utilises a single metric which only assesses a component, not all important sources, of performance. The design of the metric distorts portfolio decision-making. It is a metric which has two weak statistical properties: (1) a reasonably high likelihood of failing to identify poor funds; and (2) a reasonable likelihood of falsely identifying good funds as poor.

Further, the performance test has a range of unintended consequences, which can be costly. As one example, our research identifies the constraints on portfolio management created by the performance test to be significant. We estimate the opportunity cost of these constraints to exceed the identified benefits of the YFYS reforms.

This doesn't mean that the performance test should be removed. We reiterate our view that consumers deserve the protection of a high-quality, effective performance test. A well designed test could capture many of the identified benefits but incur fewer costs associated with unintended consequences.

We believe we have a reasonable grasp of the regulatory environment at the time of the Productivity Commission's work and the need for a bright lines test.

However, we believe it is appropriate to plan for a transition of performance testing from a legislative process (bright lines) to a regulatory process, housed by APRA. This will improve the quality of the test and ensure it has evergreen properties which account for developments in policy, industry and markets. APRA can apply multiple metrics and account for qualitative considerations. The assessment can appropriately account for the nuances of different product categories (MySuper, TDP's and Choice). The burden of proof could be structured to prevent legal challenges. We recommend that this transition take place in three years. Announcing it now will allow plenty of planning time.

The trap with analysing the performance test is that you are drawn into a black hole of trying to develop a better performance metric. This nearly always proves to be the perfect diversion from the elephant in the room: the fact that the test doesn't qualitatively account for a range of factors which are important indicators of future outcomes: from governance structures through to investment capability and resourcing. A qualitative overview has the potential to provide substantial consumer benefits.

We provide numerous insights and pieces of research in this submission. From a performance testing perspective:

1. We provide case studies which illustrate how the present test distorts portfolio management decision-making.
2. We reference forthcoming research which suggests that many approaches to account for ESG and sustainability cannot be implemented in a sustainable manner in the current performance test format.

3. We explain how many of the shortcomings of the performance test are magnified when applied to TDP's and choice products.
4. We statistically analyse the effectiveness of the current test and consider the impact of longer timeframes and more indices.
5. We suggest a small number of benchmark changes which we would prioritise.
6. We propose an additional metric, based on portfolio outcomes, which can be used alongside the benchmark-based test to improve consumer outcomes, industry fairness, and reduce unintended consequences. We recommend the adoption of this metric.
7. We suggest a more nuanced approach to the setting of tolerances and consequences.

This is a large and complex topic. We had to draw a line in the sand on our analysis. However, we would readily undertake further research at Treasury's request (we already have a large suite of existing models to leverage). This is an important area to get right.

With respect to other components of the YFYS reform package, we believe that:

- The YourSuper Comparison Tool can be improved with some small changes, and clever technology (e.g. filtering) can be used to reduce the overwhelming experience for consumers.
- Stapling is not working as effectively as it could. Hopefully Treasury is able to obtain a full understanding of the nuances in this area and decide whether changes are required.
- BFID requires some further clarity to ensure innovation, particularly in retirement services, is not inhibited.

2. Issues

2.1. Performance test

[1. Does the measurement of actual return using strategic asset allocation affect risk taking behaviour by superannuation trustees?](#)

Due to the bright lines nature of the YFYS performance test and the severe consequences (including the impact of media) it is rational for super fund trustees to ensure an extremely high likelihood of passing the performance test.

There are two elements of the performance test mechanism which warrant explanation:

1. The rolling nature of the test means that a new performance series is created each year and the benefits of positive performance test 'buffer' are limited.
2. Funds need to ensure a sustainable investment strategy (i.e. one which won't require significant alterations through time as these can be costly).

The Conexus Institute has explored this topic in detail in a paper titled "Your Future Your Super Performance Test: Constraints and Sustainable Tracking Error". The full paper has been included in Appendix 1. We find that funds will need to continue to reduce their performance test tracking error.

Further, The Conexus Institute confidentially interviewed CIOs of ten super funds (with positive performance test buffer). The paper is included in Appendix 2. This provides anecdotes of shortened investment time horizons and concerns around investment decision making.

The strong focus on performance test outcomes means that each portfolio management activity will be scrutinised on the basis of the tracking error it generates. Broadly this means:

- The investment tracking error due to active management of assets benchmarked against any of the eleven YFYS benchmarks will translate exactly to the same amount of performance test tracking error.
- Any investment activity in investment sectors which are not well-assessed by the YFYS benchmarks creates additional performance test tracking error. This form of tracking error is due to the limited number of benchmarks in the YFYS performance test.

In Case Study 1 we illustrate the issue for one sector, high yield credit.

Case Study 1

A super fund considers investing in high yield credit as a means to diversify risk sources in the portfolio and potentially enhance returns. Due to the lower liquidity profile of the high yield bond market it is difficult and costly to invest passively. The fund believes that the best managers in this space can add value. They believe a mandated 3% tracking error against the high yield benchmark would be an optimal way to invest in this sector.

For YFYS purposes, high yield debt is benchmarked against global fixed income (the global composite bond benchmark). Both the mandate and benchmark incorporate currency hedging. The calculated tracking error of the global high yield bond benchmark against the global composite bond benchmark is 9.6%.

To summarise:

- The tracking error related to implementation is 3% (the manager's mandate).
- The unintended performance test tracking error due to the benchmarking approach is 9.6%.

If we consider the possible impacts of Case Study 1:

- Funds will be constrained in how much they can allocate to this sector, and off-benchmark sectors in general.
- The case study illustrates how it is YFYS benchmarking, rather than degree of active risk taken by investment managers, which may be the largest driver of portfolio decision making.

Which strategies generate the most performance test tracking error?

- High active risk strategies – but here super funds are, and have always been, accountable for performance.
- Sectors which are not accurately benchmarked by the YFYS performance test (i.e. off-benchmark sectors). These include:
 - Sectors which diversify portfolios (e.g. credit, alternatives, emerging market equities, commodities)
 - Sectors which may provide specific risk management properties (low or high duration bonds, inflation-linked debt, commodities, alternatives, ESG strategies)

How large is the restriction in activity? Based on the research referenced above, we believe a sustainable degree of performance test tracking error is 1%. A 10% allocation to the high yield example in Case Study 1 would absorb that entire budget leaving no ability to undertake other

activities (seek returns, diversify, manage risks) which would incur performance test tracking error.

A final comment is that different activities generate varying degrees of performance test tracking error. This represents the distortion created by the test – investment sectors become relatively attractive or unattractive for a reason (performance test tracking error) unrelated to the fundamentals of the sector.

2. Does the current set of indices used to calculate benchmark returns unintentionally distort investment decisions or reduce choice for members? If so, is there a way to adjust the benchmark indices while maintaining a clear and objective performance test?

The current set of indices distorts investment decision making. We are unsure about the impact on choice for members: it depends on whether the identical set of indices would be used were the YFYS performance test applied to single sector options.

Our response to Question (1) addresses how distortion is created. In summary:

1. The test creates a performance test tracking error mindset.
2. There is a significant degree of tracking error created by different investment sectors due to benchmarking, not the active risk taken in implementation.

The distortion created by these two factors is significant while further distortion is created by one additional factor:

- The expected return of sectors against the benchmark, although subjective, is sizable in scale.

We produced Diagram 1 and Diagram 2 to illustrate the distortion.

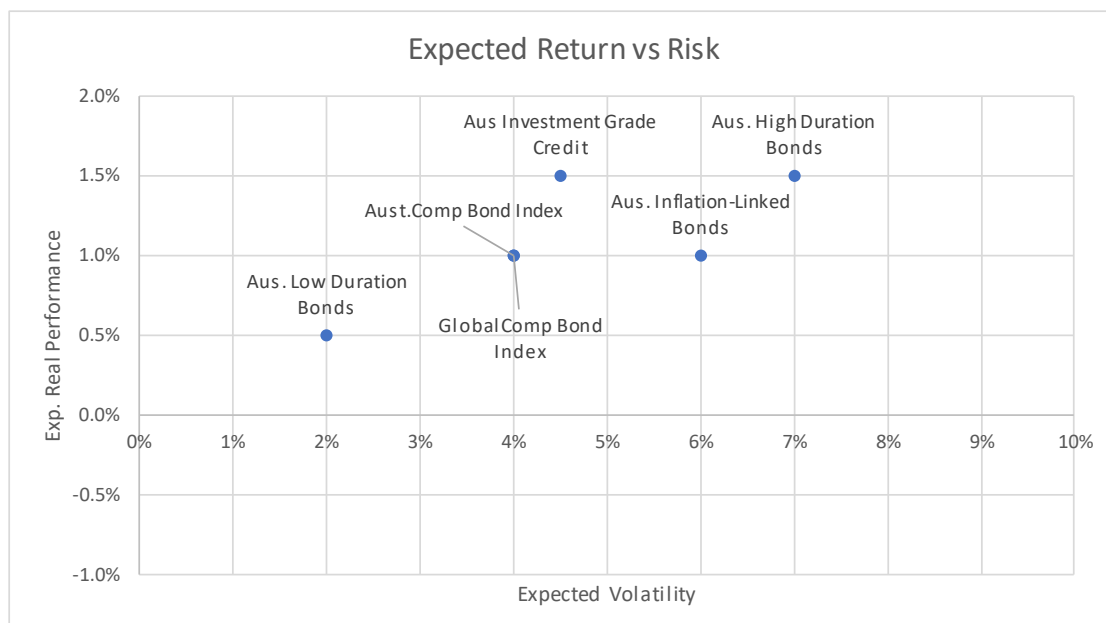


Diagram 1: Stylised expected return vs risk. Volatility is based on nominal returns for simplicity.

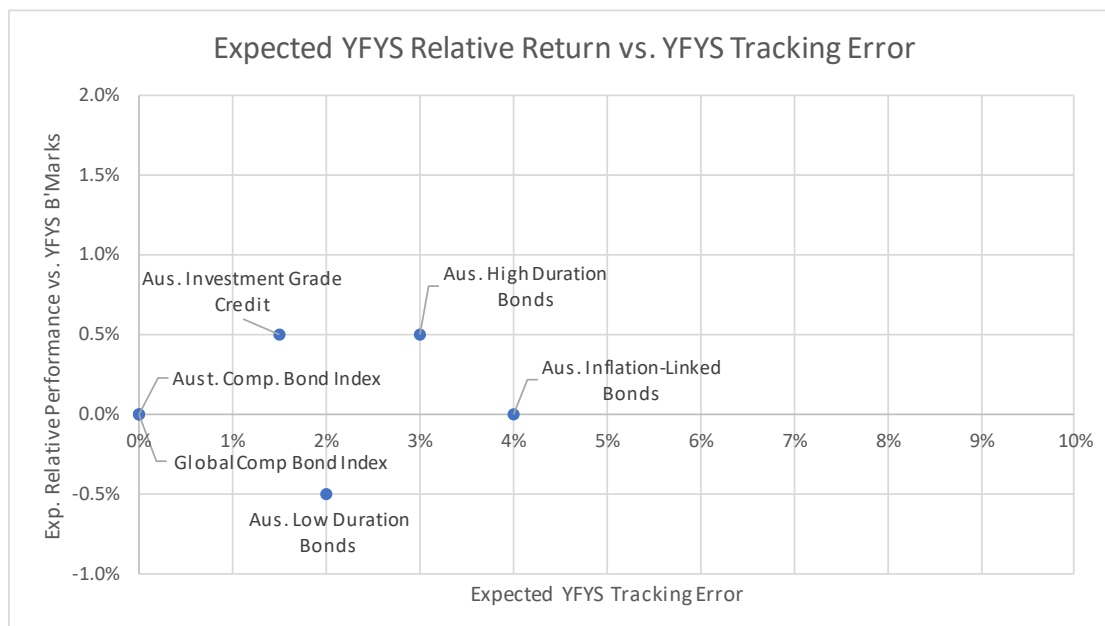


Diagram 2: Stylised relative return and risk expectations through a YFYS lens.

Comparing Diagrams 1 and 2 we make the following observations:

- In Diagram 1 there is a continuum whereby greater risk is met with an expectation of higher expected returns. This relationship falls away in the YFYS lens presented in Diagram 2.
- In Diagram 1 sectors used to benchmark outcomes are assessed on the merits of their expected return and risk properties. In Diagram 2 (the YFYS lens) these benchmarking sectors are viewed as zero expected return / expected risk.
- Lower expected return / risk sectors (e.g. Australian low duration bonds) are viewed in a balanced manner. Through the YFYS lens tracking error is agnostic to whether risk is higher or lower so low duration bonds in Diagram 2 generate positive tracking error.

There are many other sectors where distortion is significant, notably:

- Credit, where sectors are benchmarked against traditional fixed income benchmarks.
- Australian equities: through small caps and domestic private equity.
- Global equities: particularly emerging market equities and private equity.
- Unlisted infrastructure: sectors not represented in the benchmark and lower / higher risk implementation approaches.
- Unlisted property: sectors not represented in the benchmark and lower / higher risk implementation approaches.
- Alternatives: especially those strategies which diversify a 60/40 portfolio, so commodities and various hedge fund strategies.

Additionally it is important to consider ESG and sustainability related activities. Activities such as exclusions or dedicated investments can generate significant performance test tracking error. The Conexus Institute, LSEG, ASFI and RIAA are working together to explore the implications of the Your Future, Your Super performance test for portfolios which consider different ways to account for ESG, sustainability and carbon transition. The initial results suggest that the performance test tracking error created by these activities would more than absorb what is estimated as an appropriate tracking error budget for super funds.

To summarise, our most significant distortion-related concerns are strategies which are designed to manage portfolio risk and strategies designed to provide diversification benefits. We believe the size of the distortionary effects under the present test design are significant.

We believe there are two possible solutions to this problem, not necessarily exclusive:

1. Additional indices. Further indices create a more accurate measure of implementation performance at a fund level and remove much of the distortionary impacts. This improves the integrity of the performance test. Unfortunately we believe it would require 50 to 60 indices to address the variety of sector-based issue identified by industry, but much could be achieved with a smaller set of indices. The additional challenge is implementation and administration placed upon APRA and industry.
2. Additional metrics. Further metrics will shine additional light on the outcomes being produced by superannuation funds. This may enable some other portfolio construction considerations to be acknowledged, particularly other sources of return (SAA decisions) and risk.

In section (2.5) we consider two problems:

1. The possibility of adding a small number of benchmarks to the existing set.
2. The possibility of adding one additional metric to complement the benchmarks-based metric.

3. Does the calculation of actual RAFF and benchmark RAFF discourage non performance related product features that members may value (such as customer service or platform products)? If so, can this be addressed without diminishing the test's focus on performance?

Anecdotally our answer this question is yes but we have no hard evidence to offer.

We cannot see an easy solution to this challenge within the existing performance test construct. Solutions would consist of additional metrics or a qualitative overlay.

4. What are the longer-term impacts of the performance test on market dynamics and composition? How will these factors impact on long-term member outcomes?

There are many longer-term impacts of the performance test on market dynamics and composition. We trust that the broader submission process to paint the full picture. Here we detail some examples:

- Australian insurance-linked bonds: we have engaged with specialists in this area who raise concerns that, due to the high tracking error created by the benchmarking process (in this case to the Australian composite bond index), there will be a reduction in allocation to this sector which results in:
 - Reduced issuance by the AOFM
 - More expensive funding for AOFM
 - Less liquid marketplace for inflation-linked bonds
- Australian credit: due to the benchmarking process (credit is benchmarked against composite bond indices) credit becomes relatively attractive. Domestically, this may have some follow-on impacts:

- Under-priced credit risk which has potential impacts on lenders (non-super fund lenders will find the return / risk trade-off less attractive) and borrowers (potentially take higher levels of debt).
- Lower liquidity if there is little global interest in Australian debt.
- Australian private equity: due to high YFYS tracking error (private equity is benchmarked against public benchmarks) there is the possibility that Australian super funds allocate less to this sector with resulting impacts:
 - Higher cost of capital
 - Greater proportion of Australian private equity deals funded by offshore capital
- Unlisted property and infrastructure: due to the benchmarking process (a single benchmark for each sector) some biases are created:
 - A bias is created to allocate to higher risk rather than lower risk transactions. This could manifest itself in the form of higher leverage.
 - Hesitancy to allocate to opportunities in sectors not captured in the benchmarks. This may result in lower allocations by funds to emerging and niche sectors and may also result in a greater proportion of these opportunities being funded by offshore capital.

ESG, sustainability and carbon transition are special topics. From the perspective of market dynamics the greatest issues will likely centre on allocations to domestic transition infrastructure assets and new technologies.

The impact on member outcomes due to these predicted changes in market dynamics is difficult to estimate.

Separately there are many other sources of distortion which will have an impact. One is the performance test tracking error constraint which will limit the ability of funds to maximise returns, manage risk and ensure appropriate diversification. Work by The Conexus Institute estimates the cost of this impact to be \$3.1 billion per annum (refer Appendix 1).

5. Is there evidence to indicate that the notification and website publication requirements have been effective at encouraging members to consider, and switch to, alternative products? Are there ways this could be improved?

We have no informed insights.

We have consistently highlighted our concerns with the call-to-action nature of this policy. This creates a policy outcome whereby the engaged potentially benefit at the expense of the disengaged (who may be left in portfolios which aren't in best shape post-redemptions).

6. Have the consequences been effective at encouraging trustees to improve their performance or merge with better performing funds? Are there ways this could be improved?

We have no informed insight. There appears to have been a heightened level of merger activity. It is difficult to separate the performance test effect from an APRA effect (i.e. they seemed to push hard for consolidation quite hard using YFYS but also other measures such as their own fund sustainability analysis.)

7. Are the measures in place to resolve underperformance sufficient given the potential for members to be stapled to these products? How can the system best support members in underperforming products?

Our starting position is that we are far from certain that all products which fail the performance test are (1) poor funds, and (2) expected to underperform in the future given the changes they may have made. Statistical analysis presented in section (2.5) supports this position.

Consumer outcomes can be supported through decisions taken by them and above them (by super funds and regulators). In terms of supporting consumers to make decisions:

- Low financial literacy which is connected to poor financial decision-making¹. This is a difficult problem (common globally) which would take many years to address.
- Greater financial decision-making assistance, which is partly being addressed by the Quality of Advice Review (we say partly as we believe additional Government services could be part of a more holistic solution and this sits outside the scope of the Review).
- A more effective YourSuper Comparison Tool with improved disclosure accompanied by more tailored information.

In terms of structural support mechanisms:

- Effective marketplace competition. For various reasons we think competition is not as effective as it could be. We are concerned that improvements may not be forthcoming – it is a difficult problem as identified by the Productivity Commission.
- Strong, effective engagement between APRA and underperforming funds to ensure that shortcomings are addressed and that funds with structural weaknesses exit the industry. We believe this is the area where most impact is possible, and it appears that APRA is highly active with strong messages for industry while developing further areas of assessment (e.g. fund sustainability analysis).

Overall we call for greater government support for financial decision-making assistance.

8. Are there any significant issues to be expected when the test is extended to TDPs? If so, how could these issues be addressed?

The three YFYS performance test-related issues we consider most concerning for trustees (managing risk to member outcomes, achieving diversification, and generating active returns) are all magnified in a TDP environment. The reason is that for both lower and higher risk (relative to balanced funds) portfolios the opportunity set becomes smaller (a smaller number of growth investment opportunities for higher risk options or, similarly, a smaller set of defensive

¹ See for instance “[Transforming Financial ‘Advice’ Report](#)” by CoreData, which found that only 45% of surveyed Australians correctly answered all six basic financial literacy questions.

investment opportunities for lower risk options). We consider the example of conservative options in Case Study 2.

Case Study 2

It appeared commonplace for trustees of conservative options to exercise their views on the appropriate risk position within investment sectors, rather than simply the allocation between defensive sectors. This led many funds have shorter duration positions within their fixed income portfolios and more conservatively structured equity portfolios (e.g. through strategies such as low volatility equities).

Through the period up to the end of 2021 this strategy would have underperformed relative to the YFYS benchmarks. However, in the first six months of 2022 both strategies performed well against the benchmarks.

Dedicated SRI options, implemented through exclusions, are also a special case, which we detail in Case Study 3.

Case Study 3

SRI (socially-responsible investment) options commonly populate their stated ambition to managers through an exclusions-based strategy, meaning they exclude certain sectors. Carbon-intensive energy sources are a common exclusion.

Through the first six months of 2022 these strategies experienced sharply poor performance relative to the YFYS benchmarks as carbon-intensive energy stocks rose sharply in response to Russia's invasion of Ukraine.

A further issue is how lifecycle options can be implemented if a component TDP strategy fails the performance test. This could create the scenario whereby, for business stability reasons, lifecycle strategies are viewed as riskier to operate than constant risk default strategies.

If the test is to be applied to TDP products (we consider that there is an argument that they are choice products and don't need to be subject to the YFYS performance test, but respect that there are alternative views), then the suggestions we make in response to Question (2) and further explored in section 2.5, are relevant.

Further, we believe that tolerances and consequences should be re-visited. Specifically:

- It could be reasonable to allow slightly broader tolerances to acknowledge that these options are more concentrated in investment sectors and often involve trustee judgement over the best way to interpret the mandate of the option. They are also, at least part-way, choice options.
- It could be reasonable that the consequences be limited to notification, rather than product restrictions, again reflecting on the choice nature of these products.

9. What would be the impact of extending the current performance test to other Choice products (such as single sector or retirement products)? How could any issues be addressed?

To address this question we separate single sector from retirement products.

Single sector options

The current range of benchmarks will leave some products not being able to be adequately benchmarked. A possible solution for this would be to permit the YFYS benchmark to be the benchmark used in the PDS for the product.

Further, some single sector options run larger degrees of active risk which leads to a higher likelihood of false positives. Setting appropriate tolerances may assist here, and there is an argument that tolerances should be based on the tracking error of the product. We consider this further in section (2.5).

Generally the observations we made on tolerances and consequences in Question (8) apply here.

Retirement products

Most industry professionals believe that it is reasonable for consumers, especially those in default options, to be protected by a quality performance test. Yet anecdotally we observe a large industry pushback against extending the test to retirement products. We interpret this to be a reflection on the quality of the test rather than the lack of desire to protect consumers.

It is important to clarify that the most relevant category of product to apply the performance test to is account-based pensions. Some funds have highlighted the greater focus on risk management that takes place in these strategies. That argument is fair given the shortcomings of the current test design. But not many super funds run dedicated retirement investment strategies (the strategy is identical to the accumulation phase) meaning that the investment strategy would be tested if the test was extended to trustee-directed products.

2.2. YourSuper comparison tool

10. Does the comparison tool adequately inform members and prompt a behavioural response? Is the tool effective at informing new employees of their options when entering the workforce, including those who do not have an existing superannuation account?

We refer to the work undertaken by the Actuaries Institute on this topic². Our most significant concerns are:

1. The absence of any risk-based information to assist members choose an option appropriate to them.
2. No information relating to insurance costs and features.

We acknowledge that some consumer groups have flagged that many consumers feel overwhelmed by the amount of information already provided. We see this argument as having no easy solution, either:

1. Let the challenge of selecting an appropriate super option be overly simplified, which may lead to a poorly informed decision.

² [Actuaries Institute submission to ATO on YourSuper Comparison Tool.](#)

2. Provide an appropriate degree of information to support an informed decision, but risk consumers being overwhelmed.

We believe the second option is superior. Some of the challenges can be ameliorated by technology that better engages with consumers and funnels down the list of suggested funds.

11. To what extent would altered or additional metrics, or improved functionality, make the tool more effective while ensuring it remains simple and clear? What more can be done to ensure that new employees are able to choose high-performing superannuation product that are appropriate for their needs?

See our response to Question (10).

12. As the test is applied to more superannuation products, should the comparison tool also be extended? Considering the volume and complexity of Choice products, how could the tool be extended in a way that is meaningful and digestible to members?

It is a poor situation that basic product comparison information is not available to consumers at present. In that regard it makes sense for this information to be available. Greater use of filters may help to reduce the number of products to a more manageable list. There should be more disclosure that past performance may not be a good predictor of future performance.

2.3. Stapling

14. To what extent are employers putting into practice processes to seek stapled fund details from the ATO? How has the implementation of stapling changed onboarding, software and payroll processes for new employees?

Our research into this area³ has focused on understanding, but not quantifying the effectiveness of the stapling mechanism. In short, the stapling mechanism is working, but is only addressing a part of the problem, and it appears that the portion being addressed may be shrinking.

Prior to YFYS, there were two sources of multiple account creation. One was where the new employee didn't complete the Superannuation Standard Choice Form (SSCF), whereby their superannuation guarantee payments were placed into a new account with the employer's default fund.

The other was where a member elected a fund but took no action to consolidate accounts. Whether this second situation should be considered an unintentional creation of multiple accounts is debatable. I suspect it is unintentional for the majority of cases as most consumers would be unaware of the need to consolidate. The YFYS stapling measure was designed to address the first problem but not the second.

³ Our findings were published in Investment Magazine: "[Ineffective YFYS stapling measure should be reviewed](#)"

So despite the YFYS stapling mechanism, one source of unintended multiple account creation remains, and that channel is large and may be growing. This comes about if an employee doesn't complete the SSCF, an administrative workload is created for employers. It is administratively more convenient for employers to encourage employees to make a choice.

New employees are likely to choose the default fund for a variety of reasons including that it is easier than having to find the account details of their existing fund; the new employee doesn't want to "rock the boat" by creating an extra administrative burden for the employer by choosing a different super fund and the employee may interpret the default as an implicit endorsement by the employer.

In summary, it appears employers are encouraging new employees to complete the SSCF and employees are choosing the default offering. While the policy to protect consumers who don't complete the SSCF is in place, there remains a likely large and potentially growing number of multiple accounts being created, in most cases, unintentionally.

If the policy intention is to eliminate the unintended creation of multiple accounts, then a further policy response is required. The most obvious solution is some sort of account consolidation mechanism. It could be a tick-a-box on the new employment form. It would likely be more successful if account consolidation was automatic unless an opt-out box was ticked. However there are a range of issues related to personal details, verification and the risk of fraud which need to be managed.

15. Are there any barriers in the current framework to achieve the intent of the stapling reform?

Captured in answer to (14).

16. What is the actual, or likely, impact of stapling on insurance coverage?

We envisage two consequences of the outcomes identified in (14):

1. Potentially move to occupation-relevant insurance cover as employees auto-enrol into employer default options – we consider this a good outcome.
2. Potential for multiple insurance accounts if account consolidation isn't taking place – a poor outcome.

2.4. Best financial interests duty

17. To what extent has the BFID required trustees to change their processes and procedures? Has this caused any unintended consequences or impacted member outcomes in any way?

No insight.

18. Are there certain types of expenditure or activity that trustees are particularly concerned about being able to prove compliance with the BFID in respect of? Why is it difficult to demonstrate compliance? Should there be a materiality threshold?

The specific issue we explore here is retirement. We refer to the case study provided in the Exposure Draft Explanatory Materials⁴:

Example 1.1- Expenditure not in the best financial interests of beneficiaries

Yellow Super has decided to spend an amount of beneficiaries' funds in wellbeing and counselling services due to its preference for providing beneficiaries with a holistic retirement experience. While beneficiaries derive some benefits from these services, they are not financial benefits and offering the services comes at financial cost to the fund. This expenditure is unlikely to be in the best financial interests of beneficiaries.

When we extend this to retirement and the range of services a fund may consider for their members it becomes confusing. For instance:

- Household budgeting services
- Longevity assessment services
- Aged care assistance
- Connections into community groups

Each of these areas can contribute to a better retirement experience and, some more than others, a better retirement financial experience.

Additional challenges relate to the purpose of capital including ESG and sustainability, and nation-building investments.

Overall, further clarity would be beneficial. At a policy level there appears inconsistency around whether a super fund is predominantly a product provider or a 'benevolent fiduciary'. BFID needs to align with this definition. There has been a longstanding call for a purpose of super to be enshrined. We support these calls.

19. Is the reverse onus of proof the most appropriate way to achieve the objective of improving member outcomes?

No insight.

2.5. Additional issues

This section contains additional thoughts and analysis.

2.5.1. Taxation is difficult to incorporate with accuracy

⁴ Treasury Laws Amendment (Measures for a Later Sitting) Bill 2020: Best Financial Interests Obligation – Exposure Draft Explanatory Materials.

The present YFYS performance test design accounts for tax rates which are a reasonable proxy for long-term outcomes. However over shorter-time periods there can be noise. The test is administered as a compounding of short-term calculations to account for changes in SAA.

There is no easy solution to this problem, and it highlights the difficulty of (1) calculating a highly accurate performance assessment, and (2) applying it in a bright lines manner.

2.5.2. False positives and setting appropriate tolerance levels

We present some analysis on the potential for assessment errors. Supporting models for all calculations in this section are available [here](#). **The Conexus Institute would be happy to explain how the models work and undertake additional analysis.**

Test 1: Pure YFYS performance test

We ignore other sources of performance (e.g. from strategic asset allocation) and ignore other flaws of the performance test (e.g. it ignores risk outcomes). We focus purely on the performance test in a controlled environment and assess the ability of the test to accurately identify poor funds (against the performance test criteria) and falsely identify good funds as poor. Throughout this section the focus is on identifying the possibility for mis-identification due to the degree of randomness in investment outcomes at the respective tracking error levels.

Result 1: Effectiveness: probability of failing to identify a ‘poor’ fund in a controlled testing environment. We consider funds with different inherent assumptions relating to expected return and performance test tracking error.

Expected return	-0.5%	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%
	-1.0%	0.2%	7.9%	17.3%	24.0%	28.6%	31.9%
		0.5%	1.0%	1.5%	2.0%	2.5%	3.0%
				YFYS tracking error			

Result 2: False positives: probability of mistakenly identifying a ‘good’ fund as ‘poor’. We consider funds with different inherent assumptions relating to expected return and performance test tracking error.

Expected return	1.5%	0.0%	0.0%	0.0%	0.2%	1.2%	3.0%
	1.0%	0.0%	0.0%	0.2%	1.7%	4.5%	7.9%
	0.5%	0.0%	0.2%	3.0%	7.9%	12.9%	17.3%
	0.0%	0.2%	7.9%	17.3%	24.0%	28.6%	31.9%
		0.5%	1.0%	1.5%	2.0%	2.5%	3.0%
				YFYS tracking error			

Note in both results we have shaded the column corresponding with 2.5% performance test tracking error. This is our proxy for the level many funds seem to be operating at (insight gained through confidential interviews with super fund CIO’s – see Appendix 2).

What we can identify from these results is that the performance test is not overly effective at identifying inherently poor funds (Result 1). There is also a likelihood of reasonable and good performing funds being mistakenly identified as poor performers. This is all a function of the test settings: timeframes, tolerance levels and how they interact with the degree of performance test tracking error.

This provides motivation to revisit tolerances. It is a challenging topic. Setting a lower (larger negative) tolerance reduces the level of consumer protection, increases the likelihood of not identifying poor performers, but reduces the likelihood of false positives... a difficult trade-off. We haven't done this analysis as we don't consider it to be palatable (but could do so on request).

This analysis is also relevant when it comes to the single sector option discussion in Question (9). The range of tracking error levels is likely to be quite large making the setting of an appropriate tolerance level a difficult exercise.

One possible solution is the use of tolerance levels appropriate to the observed performance test tracking error of the product. This creates more targeted protection for consumers and fairer outcomes (fewer false positives) for super funds. This may be particularly relevant to choice products where the dispersion in active risk targets is significant.

Test 2: Changing the timeframe of the performance test

We use the same settings as per Test 1 except and consider performance test tracking error to be 2.5%. We consider the impact of altering the timeframe of the performance test. We account for different levels of expected relative performance and consider varying timeframes for the test.

Result 3: Effectiveness: probability of failing to identify a 'poor' fund in a controlled testing environment across different timeframes.

Expected return	-0.5%	50.0%	50.0%	50.0%	50.0%	50.0%
	-1.0%	31.2%	28.6%	26.4%	24.4%	18.6%
		6yrs	8yrs	10yrs	12yrs	20yrs
		Test timeframe				

Result 4: False positives: probability of mistakenly identifying a 'good' fund as 'poor' across different timeframes.

Expected return	1.5%	2.5%	1.2%	0.6%	0.3%	0.0%
	1.0%	7.1%	4.5%	2.9%	1.9%	0.4%
	0.5%	16.4%	12.9%	10.3%	8.3%	3.7%
	0.0%	31.2%	28.6%	26.4%	24.4%	18.6%
		6yrs	8yrs	10yrs	12yrs	20yrs
		Test timeframe				

We can observe that increasing the timeframe of the performance test has benefits in terms of a higher likelihood of identifying very poor performers (but no impact on poor performers), as well as reducing the likelihood of mis-identifying a good fund as poor. However, we consider the degree of improvement in likelihood to be modest. Any changes would have to be balanced against the extension of the impairment period (and associated opportunity cost) that funds with poor past performance would face.

Test 3: Acknowledging tracking error noise due to limited number of benchmarks

As detailed in responses to Questions (1) and (2) with a specific example in Case Study 1, the benchmarking approach creates a significant degree of performance test tracking error which isn't what industry would traditionally identify as tracking error. Our aim is to explore the extent to which this impairs the ability to accurately identify good and poor performers.

To do this we assume a base case whereby funds take 2.5% performance test tracking. We assume that this is constructed as follows:

‘True’ tracking error + additional tracking error due to limited benchmarks

We estimate the initial contribution from each source of tracking error (in the right hand column) and then consider alternative degrees of tracking error due to the benchmarking approach. Simply, more benchmarks lead to less unintended tracking error.

Result 5: Impact on effectiveness: probability of failing to identify a ‘poor’ fund in a controlled testing environment as test noise is reduced through incorporation of more benchmarks.

Expected return	-0.5%	50.0%	50.0%	50.0%	
	-1.0%	17.3%	21.6%	28.6%	
		1.5%	1.5%	1.5%	'True' Tracking Error
		0.0%	1.0%	2.0%	Due to YFYS Benchmarks
		1.8%	1.8%	2.5%	Total YFYS Tracking Error
		YFYS tracking error			

Result 6: Impact on false positives: probability of mistakenly identifying a ‘good’ fund as ‘poor’ as test noise reduced through incorporation of more benchmarks.

Expected return	1.5%	0.0%	0.1%	1.2%	
	1.0%	0.2%	0.9%	4.5%	
	0.5%	3.0%	5.8%	12.9%	
	0.0%	17.3%	21.6%	28.6%	
		1.5%	1.5%	1.5%	'True' Tracking Error
		0.0%	1.0%	2.0%	Due to YFYS Benchmarks
		1.5%	1.8%	2.5%	Total YFYS Tracking Error
		YFYS tracking error			

The results effectively explore the benefits of adding further indices. The first column reflects the addition of a large number of indices, effectively removing all benchmarking noise. The second column reflects the addition of some indices designed to reduce the largest sources of tracking error. The final column represents our base case scenario.

Broadly, adding indices helps reduce performance test noise and makes the test more effective and fairer. However, adding further indices is not a panacea. The issue identified in the previous section, having a tolerance level appropriate to the tracking error of the fund and the overall timeframe of the test, remains important.

(Note that we approximate total tracking error by taking the square root of the sum of the squares of each tracking error source. This means that we assume the two sources of risk are independent).

Test 4: Acknowledging performance due to SAA

One of the primary criticisms of the YFYS performance test has been that it focuses on one source of performance (implementation) and ignores another important source of performance, the outcome attributable to SAA decisions. It is important to understand the impact of this omission on the ability of the test to identify ‘good’ and ‘poor’ funds. Here we have expanded our definition of ‘good’ and ‘poor’ funds to not just be based on implementation but asset allocation activities as well. We assume that funds take 2.5% performance test tracking error and assume another 2% of performance volatility is created through SAA activities.

Result 7: Impact on effectiveness: probability of failing to identify a ‘poor’ fund in a controlled testing environment where expected returns from implementation and SAA activities are acknowledged.

Expected return	1.0%	13.4%								Poor: Exp. Ret = -0.5%
from SAA	0.5%	13.4%	28.3%							Very Poor: Exp. Ret = -1%
Decisions	0.0%		28.3%	49.7%						
	-0.5%			49.7%	71.6%					
	-1.0%				71.6%	86.4%				
	-1.5%					86.4%	95.6%			
		-1.5%	-1.0%	-0.5%	0.0%	0.5%	1.0%			
										Implementation Performance

Result 8: Impact on false positives: probability of mistakenly identifying a ‘good’ fund as ‘poor’ in a controlled testing environment where expected returns from implementation and SAA activities are acknowledged.

Expected return	1.5%	71.4%	50.0%							OK: Exp. Ret = 0%
from SAA	1.0%	71.4%	50.0%	28.6%						Good: Exp. Ret = 0.5%
Decisions	0.5%		50.0%	28.6%	12.9%					Very Good: Exp. Ret = 1%
	0.0%			28.6%	12.9%	4.5%				
	-0.5%				12.9%	4.5%	1.2%			
	-1.0%					4.5%	1.2%			
		-1.0%	-0.5%	0.0%	0.5%	1.0%	1.5%			
										Implementation Performance

The two results help illustrate the degree to which ignoring SAA effects impacts the effectiveness of the performance test. Note that the results are identical as we move down the columns of the two charts: this is because the test is agnostic to the level of expected SAA outcomes.

Result 7 demonstrates the potential for poor consumer outcomes as the present YFYS test ignores outcomes from poor SAA decision-makers. Funds with moderate implementation capabilities and poor SAA frameworks are highly unlikely to be identified by the performance test. Result 8 shows that for funds which are good asset allocators but poor implementers, the likelihood of falsely identifying them as a poor fund is high – a measure of rough justice for super funds, and unnecessary administration and anxiety for consumers.

The results in this section (Test 4) support arguments made for additional metrics to help paint the full performance picture. This would improve consumer protections.

2.5.3. ‘Evergreen’ framework

In present form the YFYS performance test is not ‘evergreen’. We use the term to describe a policy framework which is reasonably resilient to new issues and challenges which arise. Challenges which have arisen or become more prominent since the original YFYS performance test was introduced include:

- ESG and sustainability – additional volatility due to Russian invasion of Ukraine
- Faith-based funds
- Carbon transition and the global trend to portfolios which align with carbon reduction targets
- A high inflation environment

A proposed solution for some of these problems is a policy carve-out. This is not a great outcome as carve-outs can sometimes be taken advantage of and create additional administration.

Our approach is to aim for a better core framework which has more evergreen properties.

2.5.4. Proposing a medium-term transition of performance testing back into APRA

Our core view is that a legislated performance test creates many unintended consequences, some which our research estimates will have significant costs. We believe in performance assessment, but also believe it is a complex nuanced issue. The outcome needs to be better expected future performance for consumers. For these reasons we believe that performance testing should be integrated back into APRA over time. This means that performance testing transitions from being a legislative assessment to a regulatory assessment.

There are many benefits to this approach if done well. Performance testing is complex and nuanced and APRA have a larger toolkit. APRA has the ability to apply multiple metrics (they currently do this through their Heatmaps) as well as the capability to make qualitative considerations. This would greatly reduce unintended consequences. It would also allow the testing framework to evolve in response to industry and market developments.

The testing framework could better tie in with other policies. Consider the case of an underperforming super fund who alters their investment strategy. APRA could assess whether the investment governance framework matches up to the strategy changes.

Three years feels like a reasonable timeframe for this transition to take place. This would allow a pathway to assess additional resourcing requirements at APRA, and an appropriate pathway to unwind the legislation. Consideration should be given whether the burden of proof should reside with funds or with APRA.

2.5.5. Thought exercise: constrained to a small number of additional indices

We considered the exercise of only being able to consider a small number of additional indices. We weighted our selection on:

- Where the lack of indices is most distorting portfolio construction processes
- Where the existing test is missing important performance sources (e.g. we believe that many funds have benefitted) and additional indices could improve accountability
- Where a workable solution exists
- How commonplace the problem is amongst funds

On this basis we focus on the fixed income and credit spectrum and the alternatives sector.

For fixed income and credit we would advocate for six additional indices:

- Australian inflation-linked bonds
- High duration Australian government debt
- Low duration Australian government debt
- High duration global government debt
- High yield credit index
- Australian floating rate credit index

All international indices would be hedged into AUD.

For alternatives we would advocate for two benchmarks:

- Cash + 2%, applying to all portfolios with expected volatility of 5% or less

- Cash + 4%, applying to all portfolios with expected volatility greater than 5%

In practice a fund with an alternatives portfolio producing volatility of 7% would identify an allocation to both these benchmarks. We consider that reasonable.

While gaps would remain, we believe this would improve accountability and not deter activities intended to manage risks to consumer outcomes. The performance test tracking error of the alternatives portfolio would be the volatility of the implemented portfolio rather than the volatility of that portfolio relative to a benchmark of stocks and bonds. We consider that reasonable.

We left out a number of controversial areas, including (with comments for some on reasoning):

- Australian small cap equities: other areas considered more important
- Australian private equity: difficult to benchmark effectively
- Emerging markets equity: already a reasonable component in the ACWI index
- Global private equity: difficult to benchmark effectively, particularly new investments due to j-curve effect
- Tail risk protection: too difficult due to idiosyncratic implementation, limited number of funds
- Commodities: limited number of funds
- ESG indices: many different interpretations

2.5.6. Thought exercise: a single additional metric

We considered the exercise of adding a single additional performance metric to improve outcomes developed by the performance test.

Application of two metrics:

- Fail both metrics → fail the YFYS performance test.
- Don't fail both metrics → pass the YFYS performance test.

Description:

Performance versus the performance of a volatility-matched reference portfolio. We label this the "portfolio outcome test".

How it works:

The steps can be interpreted by reference to Diagram 1.

Step 1: A fund has realised (net of investment fees) return (A) and realised volatility (B) measured over 8 years.

Step 2: Construct a reference portfolio which has realised volatility equal to (B).

- The reference portfolio is a combination of two portfolios identical to those specified in the APRA Heatmaps:
 1. 'Growth' portfolio: 50% Australian shares and 50% global shares (50% hedged).
 2. 'Defensive' portfolio: 20% cash, 40% Australian bonds, 40% global bonds (hedged).

Calculate the realised performance of the reference portfolio, net of benchmark investment fees and taxes (C).

Step 3: Calculate fund performance including current administration fees relative to RAFE: (D)

Step 4: Calculate adjusted performance based on relative current administration fees:

$$(E) = (A) - (D)$$

Diagram 1 presents the case where a fund's investment performance (A) exceeds the performance of a volatility-matching reference portfolio (C), but its administration expenses are higher than benchmark RAFE (D). The final result (E) is that the fund's investment performance adjusted for administration fees still exceeds (C).

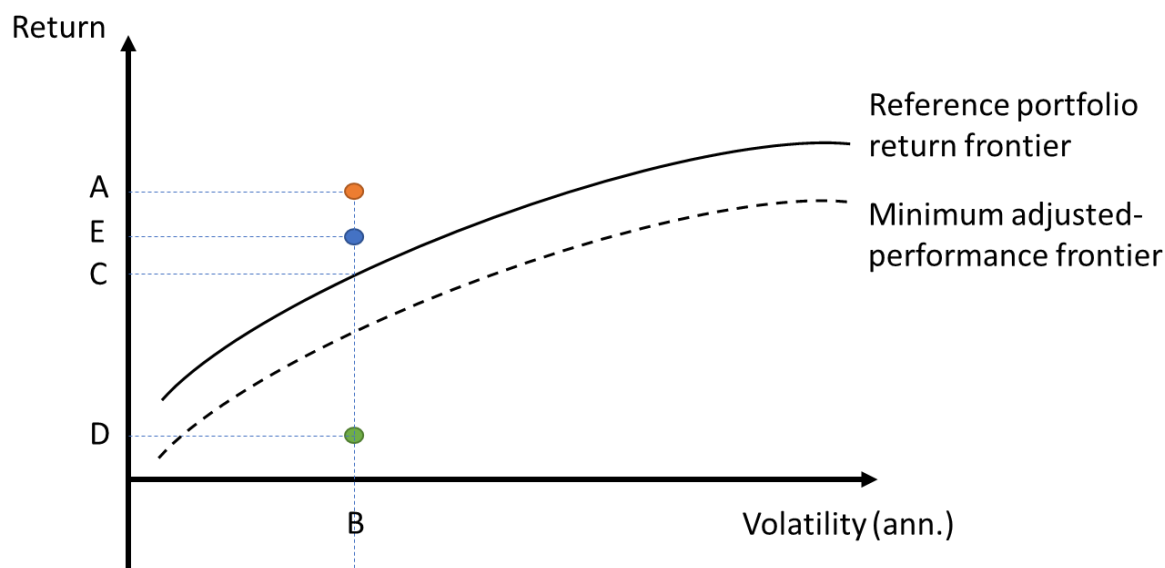


Diagram 1: Representation of alternative performance metric based referencing performance of a volatility matched reference portfolio.

All that is required to calculate the proposed metric is a time series of fund performance (recommend monthly) and performance characteristics of reference portfolio indices.

In Table 1 we self-assess the proposed metric alongside our assessment of the existing benchmarks-based approach.

	YFYS performance test (benchmark-based)	Proposed metric: portfolio outcome test
Forward-looking	✗ A backwards-looking metric	✗ A backwards-looking metric
Accounts for qualitative considerations such as process improvements	✗ No	✗ No
Accounts for administration fees	✓ Yes	✓ Yes
Accounts for implementation	✓ Yes, but numerous benchmark shortcomings	✓ Yes, implementation is captured in total performance
Acknowledges strategic asset allocation decisions	✗ No	✓ Yes, the impact of SAA decisions is captured in total performance

Deters risk management activities	✘ Yes – risk management activities can generate significant performance test tracking error.	✓/? No, the benefits of risk reduction are captured in realised volatility (acknowledging that an unrealised risk event may represent negative returns and little volatility reduction)
Deters attempts to create appropriately diversified portfolios	✘ Yes – diversification activities can generate significant performance test tracking error.	✓ No
Distorts investment decision-making	✘ Yes – benchmarking practices create performance test tracking error which does not align with member outcomes.	✓/? No, apart from relative attractiveness of unlisted assets.
Potential for gaming	✘ Yes – numerous gaming opportunities have been identified and (anecdotally) are being practiced.	✘ Yes – mainly through the processes of unlisted asset class valuations (which is subject to APRA SPG 530).
Aligns with good consumer outcomes	? Ignores SAA outcomes and portfolio risk outcomes.	✓ Yes, generally aligns with good consumer outcomes.

Table 1: Self-assessment of existing YFYS benchmark-based metric and proposed metric across multiple criteria.

Our general observations based on Table 1:

1. The proposed portfolio outcomes metric addresses many of the shortcomings identified with the existing benchmark-based metric.
2. Often, where both metrics struggle, they struggle for different reasons which means existing shortcomings are not becoming further entrenched.

There are often claims made that the volatility of unlisted assets is understated and that public market equivalents represent true insight into prices, hence the volatility of unlisted assets should be the same as listed assets. An opposing argument can be made that public market prices include an element of speculation which heightens volatility. In effect heightened volatility is the ‘cost’ of liquidity. We are not taking a view either way, rather acknowledging that there are opposing, strongly held views.

There are ways that funds which are not large users of unlisted assets can diversify their portfolios and improve their return / risk profiles. By having benchmarks embedded into the reference portfolios used in the proposed portfolio outcomes test, the assessment is localised to the fund rather than compared on a peer basis where unlisted asset exposures may then become more important.

2.5.7. Spectrum of tolerances and consequences

There is the opportunity to tailor tolerances and consequences to better suit the nature of the products and the way consumers use these products. The setting of tolerances depends on other changes being made, specifically timeframe, benchmarks and additional metrics. We assume

additional indices (per 2.5.5) and an additional metric (proposed in 2.5.6). Based on these changes our working view is presented in Table 2.

	Tolerances	Consequences
MySuper Defaults	Remain at 50bp underperformance	Existing (letter and no additional inflows).
TDP Products	Set at 50bp underperformance.	Letter to members.
Choice Products	Tolerances set to tracking error of the product.	Failure published on Government website and on fund's website.

Table 2: Proposed spectrum of tolerances and consequences, accounting for an additional performance metric and some additional benchmarks.

2.5.8. Still backwards looking...

It is worthwhile reflecting that each time there is open consultation on the YFYS performance test we are all drawn, somewhat directed, into searching for a better performance metric.

But this potentially distracts from the core issue: ensuring a reasonable prospect of good future returns for consumers.

A performance test based on past performance ignores the following current considerations:

- Quality of fund governance
- Quality of investment team
- Quality of the investment strategy
- Quality of investment capabilities and approach to implementation
- Quality of risk management

3. Conclusion

Most of our work in this submission focused on the performance test. Our conclusions are accordingly focused on this area as well.

Consumers, especially those in default options, deserve the protection of a high-quality, effective performance test. Unfortunately, our view is that the current YFYS performance test does not provide appropriate protection to consumers.

The YFYS performance test is backwards-looking and does not account for important qualitative considerations which inform future performance outcomes. The test utilises a single metric which only assesses a component, not all important sources, of performance. The design of the metric distorts portfolio decision-making. It is a metric which has two weak statistical properties: (1) a reasonably high likelihood of failing to identify poor funds; and (2) a reasonable likelihood of falsely identifying good funds as poor.

Further, the performance test has a range of unintended consequences, which can be costly. As one example, our research identifies the constraints on portfolio management created by the performance test to be significant. We estimate the opportunity cost of these constraints to exceed the identified benefits of the YFYS reforms.

This doesn't mean that the performance test should be removed. We reiterate our view that consumers deserve the protection of a high-quality, effective performance test. A well designed test could capture many of the identified benefits but incur fewer costs associated with unintended consequences.

We believe we have a reasonable grasp of the regulatory environment at the time of the Productivity Commission's work and the need for a bright lines test.

However, we believe it is appropriate to plan for a transition of performance testing from a legislative process (bright lines) to a regulatory process, housed by APRA. This will improve the quality of the test and ensure it has evergreen properties which account for developments in policy, industry and markets. APRA can apply multiple metrics and account for qualitative considerations. The assessment can appropriately account for the nuances of different product categories (MySuper, TDP's and Choice). The burden of proof could be structured to prevent legal challenges. We recommend that this transition take place in three years. Announcing it now will allow plenty of planning time.

The trap with analysing the performance test is that you are drawn into a black hole of trying to develop a better performance metric. This nearly always proves to be the perfect diversion from the elephant in the room: the fact that the test doesn't qualitatively account for a range of factors which are important indicators of future outcomes: from governance structures through to investment capability and resourcing. A qualitative overview has the potential to provide substantial consumer benefits.

We provided numerous insights and pieces of research in this submission. From a performance testing perspective:

1. We provided case studies which illustrate how the present test distorts portfolio management decision-making.
2. We referenced forthcoming research which suggests that many approaches to account for ESG and sustainability cannot be implemented in a sustainable manner in the current performance test format.
3. We explained how many of the shortcomings of the performance test are magnified when applied to TDP's and choice products.
4. We statistically analysed the effectiveness of the current test and considered the impact of longer timeframes and more indices.
5. We suggested a small number of benchmark changes which we would prioritise.
6. We proposed an additional metric, based on portfolio outcomes, which can be used alongside the benchmark-based test to improve consumer outcomes, industry fairness, and reduce unintended consequences. We recommended the adoption of this metric.
7. We suggested a more nuanced approach to the setting of tolerances and consequences.

This is a large and complex topic. We had to draw a line in the sand on our analysis. However, we would readily undertake further research at Treasury's request (we already have a large suite of existing models). This is an important area to get right.

With respect to other components of the YFYS reform package, we believe that:

- The YourSuper Comparison Tool can be improved with some small changes, and clever technology (e.g. filtering) can be used to reduce the overwhelming experience for consumers.
- Stapling is not working as effectively as it could. Hopefully Treasury is able to obtain a full understanding of the nuances in this area and decide whether changes are required.

- BFID requires some further clarity to ensure innovation, particularly in retirement services, is not inhibited.