Working Version: Review of the Your Future Your Super Performance Test

Summary Paper

20 November 2020

David Bell, Emily Barlow, Andrew Boal, Kim Bowater, Nick Callil, David Carruthers, Matthew Griffith, Clayton Sills, Tim Unger

Table of Contents

1. Introduction

The Government recently announced the Your Future Your Super (YFYS) reforms as part of the 2020 Budget. There is merit in protecting consumers with a performance test. However, it needs to be an effective performance test with limited undesirable outcomes. Unfortunately, our analysis suggests the YFYS performance test does not meet these goals: it will be ineffective at identifying poor performing funds while introducing a range of undesirable outcomes. We are concerned that the detriments of the YFYS performance test may outweigh the benefits.

To summarise our analysis:

- An investment performance metric has shortcomings: it doesn't account for the total consumer outcome (by considering product design), fails to account for changes made through time by super funds, while the evidence that past performance is a predictor of future performance (performance persistence) is modest (total fees and governance appear informers of future outcomes).

- Placing these issues aside we assessed the YFYS performance metric. The statistical effectiveness of the YFYS performance metric at identifying poor performing funds is found to be weak. The YFYS performance metric faces three major challenges: (1) timeframe (8 years may not be sufficient), (2) it focuses on one (likely minor) component of performance (implementation[1]) rather than investment performance in total, and (3) the benchmarking process generates inaccuracies which create benchmark 'noise'. Our statistical analysis reveals that, in its current form, the performance metric will be ineffective at identifying poor performers and have a high likelihood of falsely identifying good performers as poor.

- We believe the YFYS performance test will result in undesirable outcomes relating to how funds invest, may have some adverse impacts on consumers, and create a distorted industry structure (the potential for 'zombie' funds). We expect there to be a detrimental effect on both industry performance and individual consumer outcomes.

We detail a range of alternative solutions and self-assess through the lens of consumer outcomes. To summarise:

- A solution which involves regulator assessment is, as it stands, the only way to acknowledge the qualitative issues and the through-time changes made by funds. It can also incorporate metric(s).

[1] Here, implementation performance includes asset level performance assessed against ascribed sector benchmarks and the performance of tactical asset allocation calls (performance from having an asset allocation which is different from the fund's strategic target).

- Better metrics exist and a well-designed collection of metrics has advantages over a single metric.
- The YFYS performance metric can be improved. At best it can be more stable and assess all fund-types consistently. Nonetheless, it will remain statistically ineffective at assessing implementation performance, which itself is a small component of total investment performance.

We are positive about the opportunity to improve consumer outcomes. There is a great opportunity to implement an effective performance test which protects consumers from being exposed to funds which are likely to underperform in the future, whilst limiting undesirable outcomes. We are happy to share models and discuss this work and solutions in further detail.

This paper presents a summary of substantial analysis undertaken on the YFYS performance test. It is accompanied by a Detailed Paper for those who want to explore further.

2. Analysis

2.1. What drives consumer outcomes?

Superannuation is a complex system. In Diagram 1 we provide an overview through the lens of the outcomes it delivers. Investment performance is an aggregate of three components, in order of importance: (1) the level of risk, (2) asset allocation, and (3) implementation. Investment performance is delivered within a product design which may vary the risk target through time to account for expected future contributions (known as a lifecycle strategy). The superannuation product applies to a range of members with unique characteristics relating to their own income/contribution experiences. A range of possible outcomes is the result with an observed outcome effectively being one realisation of a process (approaches for assessing the range of outcomes, including case studies which incorporate lifecycle strategies are provided in the Detailed Paper).

In our view no component of performance should be ignored as each can have a significant impact on consumer outcomes and each is an active trustee decision (demonstrated in the Detailed Paper).

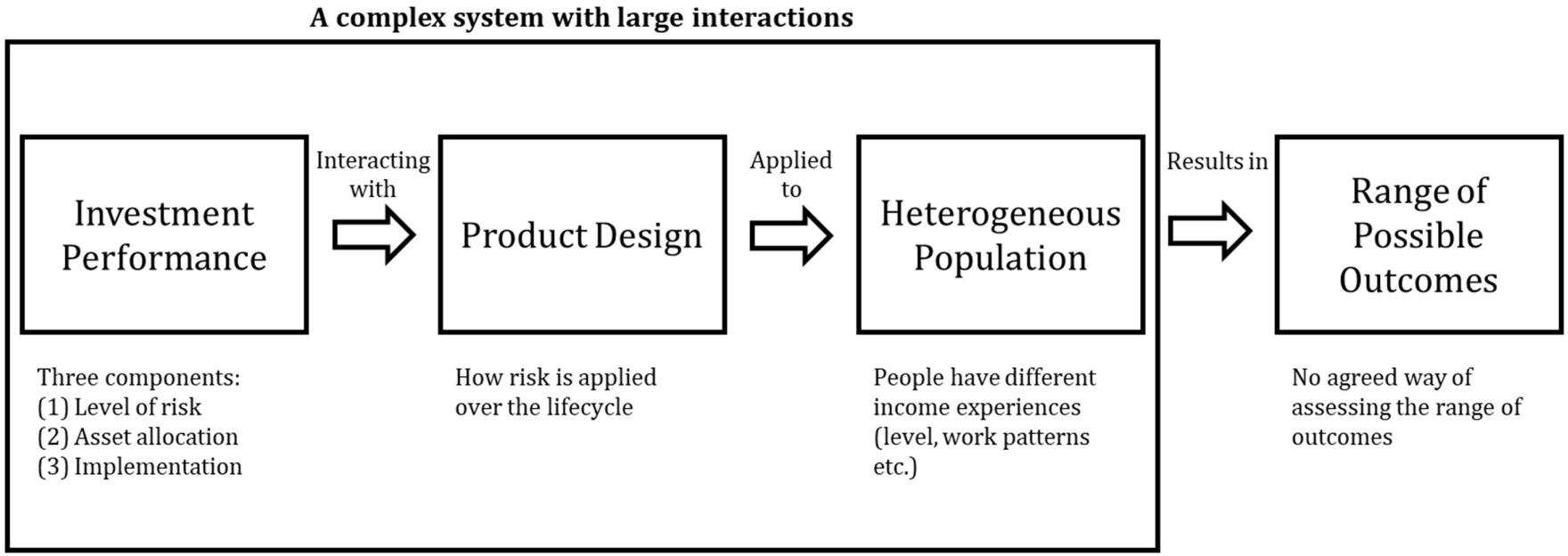

Diagram 1: Overview of superannuation system from the perspective of the outcomes it produces.

As detailed in Diagram 2, returns are not the only output from the investment process: realised experiences of variability are also an output. Here we use volatility as a descriptive measure of realised risk. The Detailed Paper explains the importance of risk reduction while highlighting that gross volatility is a poor proxy for net volatility. We advocate for an approach that recognises net risk.

Ultimately part of the YFYS performance test is to improve consumer outcomes and create an environment of enhanced accountability. In this regard it is important that the test focuses on the performance outcomes that impact consumers (total investment performance), that it acknowledges net risk rather than gross risk (to include portfolio construction in the assessment) and that it recognises total fees (as fees provide most insight into future performance). Incorporating qualitative analysis would allow the assessment of governance.

This is reflected in Diagram 2. We believe that it is important for consumers that all the outputs identified in Diagram 2 are considered in any assessment of performance.

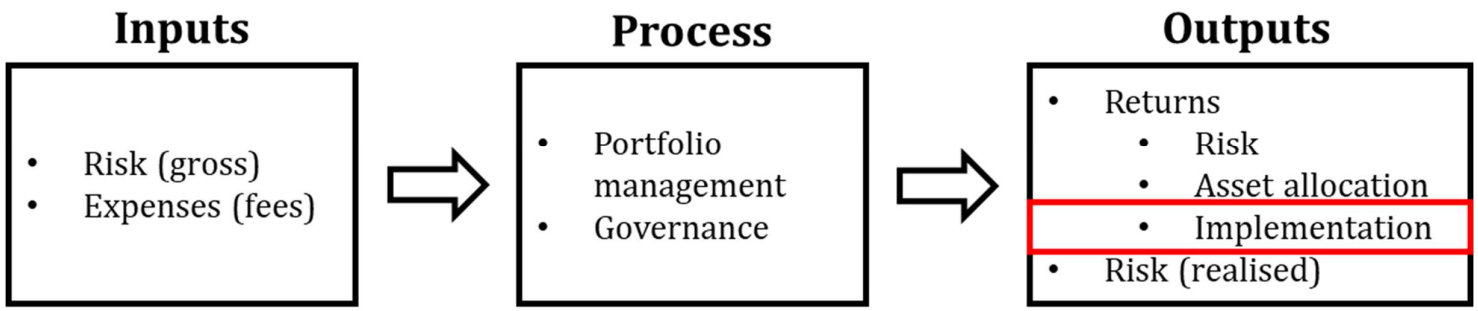

Diagram 2: Process representation of investment management. The red box reflects the focus of the YFYS performance test. Note that chosen inputs (risk and expenses) account for the assessed opportunity set (expected returns, forecast volatility, correlations etc.), fund membership characteristics (demographic profile), and fund-level constraints such as liquidity.

2.2. Persistence of investment performance

Academic research (see the Detailed Paper for full analysis) suggests that, generally, performance persistence is weak and that fee loadings and quality of governance are indicators of future performance. Active management is a side issue: there is a case that active management can improve net returns for institutional investors, however, we believe an effective performance test will increase trustee accountability around performance outcomes. Importantly, funds are not stationary: they make changes through time. This limits the ability to rely on past performance as a predictor of future performance. Given areas such as fees and quality of governance impact future performance, it is important to account for these through-time changes. Placing these issues aside we explore the effectiveness of metric-based performance tests.

2.3. The effectiveness of the YFYS performance metric at identifying 'poor' funds

We place aside the issues of total product design, persistence and non-stationarity and simply address the question: will the YFYS performance metric be effective at identifying poor performing funds? Unfortunately, we find the answer to be no. The YFYS performance metric will be ineffective at

identifying poor performers and have a high likelihood of identifying good performers as poor. Statistically, the YFYS performance metric faces three major challenges detailed in Diagram 3.

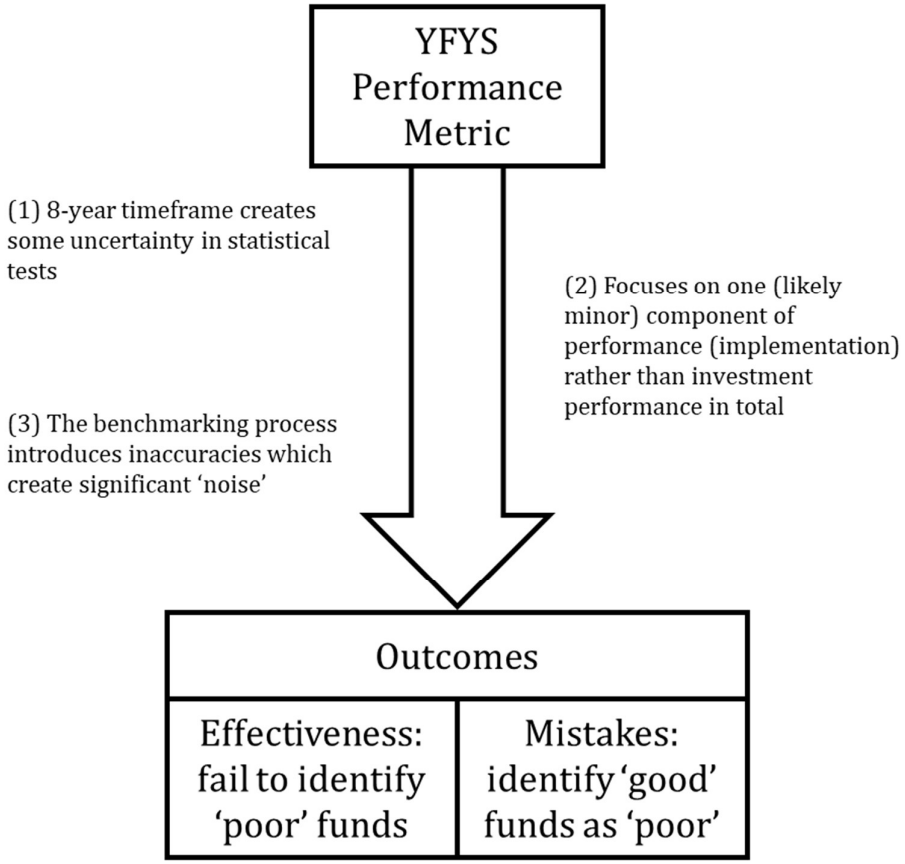

Diagram 3: Detailing concerns around the effectiveness of the YFYS performance metric.

In Diagram 3 benchmark noise is created by benchmarking sectors against indices which do not track their short-term performance. There are many sectors impacted, most notably private equity, unlisted property, unlisted infrastructure, all forms of credit, inflation-linked bonds, and the entire universe of alternative assets.

We developed a statistical test (the models for which are available) to assess the overall effectiveness of the YFYS performance metric. Our results are summarised in Table 1. Overall, we consider the effectiveness of the test to be very weak (in the context that a coin-toss would result in 50% outcome for effectiveness and mistakes).

Test	Likelihood
Effectiveness: likelihood of failing to identify a 'poor' fund as 'poor'	42% - 65%
Mistakes: likelihood of identifying a 'good' fund as 'poor'	35%

Table 1: Assessed statistical effectiveness of the YFYS performance metric. For more details see the Detailed Paper.

We found that, when we aggregate these issues, the YFYS performance test is likely to have weak effectiveness at identifying poor total investment performers across a range of different types of super funds. Even if policymakers decide to focus solely on implementation the existence of significant benchmarking noise renders the test highly unreliable.

2.4. Undesirable outcomes

We believe the YFYS performance test will result in undesirable outcomes relating to how funds invest, direct impact on consumers and industry structure. Our concerns are summarised in Table 2 with greater detail provided in the Detailed Paper.

Concern 1: How funds will invest	Concern 2: Direct impact on consumers	Concern 3: Impact on industry structure
- Distortion to portfolio management behaviour. Focus on reducing tracking error to performance benchmark and through-time performance management will make trustees more short-term focused and make a range of 'noisy' investment strategies less attractive. We expect this to increase portfolio turnover costs and reduce portfolio quality. - Dangerous incentive for funds which are well behind on the performance test to 'swing for home runs' and take high tracking error relative to benchmark. - Actively managing the performance test by taking advantage of benchmark shortcomings. - Poor alignment with portfolio management approaches such as total portfolio approach (TPA). - Deterrent to strategies which reduce risk and provide diversification. - Management of ESG risk creates more benchmark "tracking error".	- Given the low effectiveness of the test super funds may 'contest' the result with their members, creating confusion. - The YFYS performance test result may create confusion for consumers when placed alongside total performance on the YFYS Comparison Tool. - Potential for a large cohort of funds to fail the YFYS test concurrently (due to benchmarking noise), reducing system confidence. - Doesn't remove consumers from investment products with assessed high administration fees. - Penalises the heavily disengaged who may remain in a fund which becomes more impaired.	- A deterrent to consolidation as funds will be hesitant to merge with other funds which may dilute their portfolio quality, impair their inflow profile, or distract management focus. - Potential 'zombie' funds which are impaired, making them an unattractive merger partner.

- Potential for reduced investment in Australian unlisted assets. - Features of the YFYS performance test do not match up well with future portfolio management challenges.		

Table 2: Summary of undesirable outcomes likely to result from the YFYS performance test. For more details see the Detailed Paper.

3. Solutions

There is merit in providing consumer protection via a performance test. However, it needs to be an effective performance test, and one with limited undesirable outcomes. Often it is easier to critique a solution than propose a better one. We consider a range of different solutions, each of which we believe would improve consumer outcomes relative to the YFYS performance test. We begin by detailing a set of principles centred on consumer outcomes, detailed in Diagram 4, which guide our solution assessment. We distinguish between a pure investment test and a consumer investment outcome test, which incorporates administration fees.

1. Informs future underperformance

- A performance test has greatest benefit if it reliably informs future performance
- Total fees and quality of governance inform future performance
- Accounts for non-stationary challenges by accounting for the changes made by funds through time

2. Assesses total investment performance

- Acknowledges all sources of investment returns (level of risk, asset allocation and implementation)
- Appropriately recognises net risk

Diagram 4: Principles for designing a performance test which will protect consumers from experiencing poor performance in the future.

Based on our assessment (see the Detailed Paper for more granularity) we make a number of observations:

- A solution which incorporates regulator assessment has the greatest potential to be effective. It affords the opportunity to recognise total outcomes to members, assess total investment performance, and account for the through-time changes made by funds. It allows the integration of both multi-metric-based and qualitative assessment techniques (in areas such as governance). We note that the Productivity Commission suggested APRA perform a role of this nature in certain situations.

- A collection of metrics, all else equal, is superior to an individual metric. Any individual metric will have shortcomings and these can be reduced through the judicious use of additional metrics. However, metrics can only go so far: they cannot always account for changes made through time while some areas such as governance are best assessed qualitatively.

- There are alternative individual metrics which we believe are superior to the YFYS performance test. There are alternative metrics to the YFYS metric which better account for total investment outcomes delivered to consumers and a smaller degree of undesirable outcomes

- The YFYS performance metric can be improved (the Detailed Paper provides a number of recommendations). At best it can be more stable and treat all funds more fairly. It will remain statistically ineffective at assessing implementation performance. It fails to focus on total investment performance to consumers. The YFYS metric, even if improved, will create a large range of undesirable outcomes.

4. Conclusion

In this Summary Paper we condense the substantial analysis undertaken on the YFYS performance test. There is merit in protecting consumers with a performance test. However, it needs to be an effective performance test with limited undesirable outcomes. Unfortunately, our analysis suggests the YFYS performance test does not meet these goals: we have strong concerns that the YFYS performance test will be ineffective at identifying poor performing funds while introducing a range of undesirable outcomes. We are concerned that the detriments of the YFYS performance test could may outweigh the benefits.

Our concerns relate to what the metric actually assesses (a small component of total investment performance), the limited ability of backwards-looking performance measures to predict future performance, the statistically weak effectiveness of the YFYS metric for distinguishing between 'good' and 'bad' funds, and the anticipated impacts on fund behaviour, consumer outcomes and industry structure. In summary we have reservations about the effectiveness of the YFYS performance test and believe it risks significant undesirable outcomes.

Nonetheless, we are positive about the opportunity to improve consumer outcomes. There is a great opportunity to implement an effective performance test which protects consumers from being exposed to funds which are likely to underperform in the future, whilst limiting undesirable outcomes.

We detail some solutions based on principles relating to consumer outcomes, namely a measure which provides some insight into future performance, and one which assesses total investment performance. This leads us to explore a range of solutions. We believe a solution which incorporates regulator assessment is best, assess that a collection of metrics will be more effective than an individual metric, and identify that there are other metrics which could perform more effectively and with less undesirable outcomes than the YFYS performance metric. Finally, we note that a range of modifications could improve the YFYS performance metric, but it would remain ineffective relative to alternative solutions.

The Detailed Paper provides far greater detail and references. We are happy to share models and discuss this work and solutions in further detail.