CMSRJ

Council for Medical Schemes Research Journal
Vol. 1 (1) · 2026

Inaugural Issue: CMS Research Symposium



Evidence for a fairer, more sustainable health-financing system

11	4	49
ARTICLES	RESEARCH STREAMS	AUTHORS

COUNCIL FOR MEDICAL SCHEMES – PROTECTING THE INTERESTS OF THE BENEFICIARIES
ISBN 978-1-0483-6688-4 (print)





ABOUT THE JOURNAL

EVIDENCE FOR A FAIRER, MORE SUSTAINABLE HEALTH-FINANCING SYSTEM

PUBLISHER

The Council for Medical Schemes Research Journal (CMSRJ) is published by the Council for Medical Schemes (CMS), the statutory regulator of the medical schemes industry under the Medical Schemes Act 131 of 1998. Editorial office: Block A, Eco Glades 2 Office Park, 420 Witch-Hazel Avenue, Eco Park, Centurion 0157, South Africa.

AIMS AND SCOPE

CMSRJ is an annual, peer-reviewed, open-access journal publishing original research, case studies, statistical releases, reviews and policy analysis that advance the Council's mandate to protect beneficiaries and sustain private health financing. Its interests span regulation and supervision; governance and beneficiary protection; benefit design and financial protection; industry statistics and health systems; and strategic purchasing and the path to universal health coverage.

Identifiers: ISBN 978-1-0483-6688-4 (print edition of this issue). This issue: Volume 1, Issue 1 (2026): Inaugural Issue, CMS Research Symposium. Published annually.

Peer review and ethics: All manuscripts undergo double-anonymised review by at least two independent referees. Authors declare conflicts of interest, funding and, where relevant, ethics approval and an AI-use statement. Studies using human data confirm appropriate consent and ethical clearance.

Disclaimer: Views expressed are those of the authors and do not necessarily represent the official position of the Council for Medical Schemes. Publication does not constitute endorsement of any scheme, product or commercial programme named by the authors.

CMSRJ EDITORIAL TEAM

EVIDENCE FOR A FAIRER, MORE SUSTAINABLE HEALTH-FINANCING SYSTEM



CMSRJ EDITORIAL TEAM

The journal is stewarded by an editorial team from the Council for Medical Schemes.

EXTERNAL REVIEWERS

Prof Plaxcedes Chiwire (Western Cape Government: Health and Wellness; University of Cape Town: Health Economics)
Prof Edmore Marinda (University of Johannesburg)
Dr Edith Madela-Mntla (University of Pretoria)
Prof Olufunke Alaba (University of Cape Town)

CMS INTERNAL REVIEWERS

Dr Chuma Makunga
Dr Keolophile Makgamathe
Mr Kgothatso Nkoana
Mr Selby Masekela
Mr Phakamile Nkomo
Ms Irmin Segeel
Ms Ronelle Ferreira

EDITOR IN CHIEF

Dr Michael Willie

MANAGING EDITORS:

Dr Samantha Iyaloo
Mr Stephen Monamodi

ASSOCIATE EDITORS

Mr Percy Daames
Mr Joe Modise
Ms Busi Khuzwayo

PUBLISHED BY

Council for Medical Schemes

MANAGING OFFICE:

editor@medicalschemes.co.za

Reviewer acknowledgements: CMSRJ operates a double-anonymised peer-review process. The editors gratefully acknowledge the reviewers whose careful, constructive and timely assessments upheld the scholarly standards of this inaugural issue. Their contribution, given freely and in confidence, is central to the integrity of the journal.



ADDRESS

EVIDENCE FOR A FAIRER, MORE SUSTAINABLE HEALTH-FINANCING SYSTEM



DR MUSA GUMEDE · CHIEF EXECUTIVE AND REGISTRAR, COUNCIL FOR MEDICAL SCHEMES

REGULATION IS ONLY AS GOOD AS THE EVIDENCE BENEATH IT

The Council for Medical Schemes protects the interests of beneficiaries and safeguards the soundness of a private health-financing system that covers millions of South Africans. That duty is discharged best when it is informed by evidence rather than assertion. Every regulatory judgement the Council makes, whether on the benefits that schemes must fund, the rules and contributions it approves, or the conduct of the markets that supply members, rests on a view of what works, for whom and at what cost. Where that view goes untested, regulation risks protecting established practice rather than beneficiaries.

South Africa's private health sector has never been short of opinions about what ails it. What it has lacked is evidence that is shared, published and open to challenge. Much of the most valuable analysis in the sector is held within schemes, administrators and managed care organisations, where it informs individual decisions but seldom reaches the public record and cannot be examined, replicated or contested by others. The CMSRJ exists to change that. By giving research presented at the CMS Research Symposium a permanent, peer-reviewed home, it allows the evidence on which policy is argued to be read by every party to the debate.

This inaugural issue gathers the research presented at the 2026 Symposium: eight full research articles and three abstracts, arranged in four research streams. Its questions are the Council's questions: how continuity and coordination change the cost and quality of care; how benefits and health technologies should be assessed and funded; how value, rather than volume, can be purchased; and how the markets that supply our members can be kept fair and transparent. Read together, the studies trace a single arc, away from fragmented, hospital-centric, fee-for-service care and towards coordinated primary care, evidence-based benefit design, transparent health technology assessment, and strategic, data-driven purchasing. That direction matters well beyond the private sector, because the institutions, methods and data built now will shape South Africa's path to universal health coverage. I commend the authors for work that is rigorous, practical and unafraid of difficult findings.

The Council will judge this evidence, as it must judge all evidence, by what it means for beneficiaries. Lower claims costs and better clinical markers are important, but they are not the whole of the matter. A regulator must also know whether savings to a scheme have been shifted to members as out-of-pocket payments, and whether members can reach the care they need when they need it. I therefore encourage researchers to treat financial protection, access and members' own experience of care as outcomes in their own right. The foreword that follows considers how far the present issue already does so.

A journal published by a regulator carries particular obligations. The CMSRJ will hold itself, as well as its authors, to the standard it asks of the sector. It will publish rigorous work on regulation, benefit design, purchasing and market conduct, whether that work comes from universities, schemes, providers or the Council itself, and it will review submissions without fear or favour through double-anonymised peer review. Null and negative results are as welcome as positive ones, because a regulator learns as much from what does not work as from what does. The Council convenes this evidence and welcomes the debate it will provoke, but publication does not constitute endorsement of any scheme, product or commercial programme named by the authors.

The findings belong to their authors; the questions belong to all of us.

FOREWORD

EVIDENCE FOR A FAIRER, MORE SUSTAINABLE HEALTH-FINANCING SYSTEM





DR MICHAEL WILLIE · EXECUTIVE: POLICY, RESEARCH AND MONITORING

EVIDENCE WE CAN ACT ON, AND EVIDENCE WE MUST TEST

The Chief Executive's address sets out why the Council needs evidence that is shared and open to challenge, and the direction in which the studies in this issue point. The task of this foreword is narrower and more critical: to ask how far that direction is supported by the evidence presented, and at what cost and to whom.

The contributions are not all of one kind, and they should not be read as though they were. Some are large analyses of claims data, while others are simulation models, practice accounts distilled from a decade of implementation, or policy analyses, and each design supports conclusions of a different strength. The discussion that follows takes the four research streams in turn before considering the issue as a whole.

The case for coordinated primary care, long argued on international evidence, is now gaining empirical weight in South Africa. In an analysis of 452,793 members, most with chronic conditions, those who stayed with one GP had lower risk-adjusted costs, with fewer specialist visits, diagnostic tests and admissions, than those who moved between doctors. Because continuity was encouraged partly through a copayment for consulting a GP other than the nominated one, a regulator must also ask whether any savings to the scheme came from costs shifted to members or from care deferred by those least able to pay. That question can be answered only when out-of-pocket spending is analysed alongside scheme costs, which should become a standard feature of such evaluations. A practice account of integrated primary care hubs, which now serve more than 135,000 members in five provinces, suggests that GPs must first integrate with care coordinators, allied health professionals and local services such as mental health and hospice care before integration with specialists and hospitals can succeed. A modelling study suggests that nurse-led team practice could serve about 3.8 times as many patients as a solo GP, assuming a fee per visit about 36% lower (R350 rather than R550), while increasing the GP's net practice revenue by about 41%. These estimates assume more GP consulting time than in solo practice and depend on throughput, supervision arrangements and nurse supply that will vary between settings, so they must be tested in working practices before they can responsibly inform contracting.

Decisions about what medical schemes fund, and how those decisions are reached, need to be made in the open. The second stream turns to the prior question of what should be funded, and its first paper argues for a single, independent and transparent health technology assessment agency, with methods suited to medical devices and diagnostics rather than borrowed wholesale from medicines. The second paper describes how international guideline evidence was adapted, through GRADE adolopment and deliberation by a multistakeholder committee, to develop evidence-informed definitions of Prescribed Minimum Benefits. Adolopment is efficient, but it imports judgements about cost-effectiveness, prices and priorities made for other health systems, so the rigour of the local adaptation, more than the quality of the source guidelines, will determine whether the resulting definitions fit South African conditions. Read together, the two papers also expose an unresolved question of institutional design: how a national assessment agency, in a system moving towards universal coverage, should relate to the benefits that medical schemes are obliged to fund.

Payment should follow clinical need rather than service intensity, and medicine prices should reflect fair competition rather than market distortion. Among 9,878 adult psychiatric patients, those who shared a diagnosis followed markedly different trajectories depending on the severity of their distress at admission, and the authors conclude that diagnosis alone is insufficient to determine the most appropriate level of care. A stepped-care model that links funded care to measured severity would, however, create an incentive to record higher severity, particularly where the measure is a self-reported scale such as the DASS-21, so any payment model built on it will need independent audit of the scores that trigger it. An

evaluation covering 80,719 members reports that those enrolled in a diabetes programme were more likely to receive an annual HbA1c test (82% against 59%) and had 15% fewer hospital admissions in 2024. Enrolment, however, was voluntary and reached fewer than half of eligible members, so part of this difference may reflect who chose to join rather than what the programme achieved. In medicines, a composite index of utilisation concentration and price dispersion flagged 57% of generic markets as high-scoring in national sales data from 2014 to 2015 and 52% in scheme claims from 2020 to 2024. Because the two datasets differ in period, coverage and payment setting, that similarity need not mean similar market conditions; it may equally reflect differences in context or the sensitivity of the index itself. Until its signals have been validated against subsequent regulatory findings, the index is best treated as a screening device that directs attention rather than as evidence of misconduct.

Linked clinical, social and behavioural data can help to turn prevention from an aspiration into routine practice. A long-running practice account reports that patients with both high clinical and high social or functional risk used far more hospital bed days than those with high clinical risk alone, a burden that diagnosis-only risk models are likely to underestimate, although the underlying rates have yet to be reported. Among members whose baseline sleep habits were Poor or Medium, those whose sleep improved had in-hospital costs about 7% lower in the following year, although the 95% confidence interval includes zero. In a cohort of 28,248 people with stage 1 cancer, any recorded physical activity before diagnosis was associated with lower rates of progression and all-cause mortality, with larger reductions at higher activity levels. Because sleep and activity were captured through wearable devices, gym visits and organised events, the most engaged members are likely to differ from others in ways that no dataset fully records. Device ownership and time to exercise are also unevenly distributed by income, so incentive programmes built on such data could, whatever their intention, reward members who are already healthier and operate as a subtle form of risk selection. Read across the streams, the convergence of these findings is encouraging, but it is less independent than it first appears. Several studies draw on the same large open scheme, and several evaluate programmes designed and run by the organisations that employ their authors, so agreement among them may partly reflect shared data, shared assumptions and shared interests rather than independent confirmation. Neither feature is a reason to dismiss the work, since insiders often hold the best data and the full articles declare their conflicts of interest, but both call for care in interpretation. Most of the empirical studies are also observational, and members who consistently consult one GP, enrol in a disease programme or wear a fitness device are unlikely to resemble those who do not, a risk that the authors of the continuity study themselves acknowledge. Risk adjustment and causal inference methods can account for observed differences but not for unmeasured ones such as motivation or health literacy. The three contributions published here as abstracts should likewise be read with caution and, where they have been published in full, alongside the full articles, since an abstract cannot carry the detail needed for scrutiny.

What the issue does not yet measure deserves as much critical attention as what it does, because omissions shape policy as surely as findings do. Measured against the beneficiary outcomes identified in the Chief Executive's address, the gap is clear. Success is judged mainly through claims costs, utilisation and clinical markers, while member-reported experience, out-of-pocket spending and access to care seldom serve as primary outcomes. The evidence also comes predominantly from large schemes and their administrators, leaving members of smaller schemes and low-cost options, and the interface between private and public care, largely unexamined. What the field now needs is independent replication, comparison across schemes, and designs such as staggered roll-outs, natural experiments and, where feasible, pragmatic randomised evaluations that allow causal claims to be tested rather than assumed.

I therefore invite researchers to bring their strongest work to these pages, including studies that test or overturn the findings reported here, and invite those who disagree with what they read to set out their case in the same forum. Only through that exchange will the sector's evidence become as robust as the decisions that rest on it.

My thanks go to our authors for their candour, to our external and internal reviewers for their rigour, and to the editorial team for turning a symposium into a journal.

CONTENTS

IN THIS ISSUE: FOUR RESEARCH STREAMS

STREAM I: VALUE-BASED CARE MODELS 8

One GP: the case for primary care coordination in chronic care 8
Phumla Litlhakanyane, Claudia Pastellides, Aminda Bronkhorst, Amalia Theologides and Lara Wayburne

Implementing integrated primary care hubs in private healthcare: lessons from a decade of practice in South Africa 26
Hellen Nkwagatse and Belinda Richards

To what extent can nurse-led, team-based PHC delivery models improve cost efficiency, access and provider incentives relative to traditional solo GP models in South Africa? A shared value assessment of alternative models for primary care delivery 39
Roseanne Harris, Nondumiso Gugu Beatrice Khumalo and Sarika Besesar

STREAM II: HTA AND PRIORITY SETTING 59

Toward universal health coverage: reimagining the health technology assessment framework in South Africa 59
Mark A. Brand

Developing health technology assessment–informed Prescribed Minimum Benefit definition guidelines using an adapted GRADE adolopment approach 78
Ameer Hohlfeld, Shelley McGee, Millidhashni Reddy and Trudy D Leong

STREAM III: VALUE-BASED CONTRACTING AND PRICING REFORM 79

Beyond bed days: clinical outcomes to inform psychiatrist-led stepped care and value-based contracting in mental healthcare 79
Marisa Henderson, Juanita Swanepoel and Annamarie van Jaarsveld

A composite market distortion index for regulatory surveillance: detecting price–volume anomalies in generic pharmaceutical markets in South Africa 104
Molaudi Gaula, Alex van den Heever, Ntobeko Mpanza, Neo Kitsa, Yongama Njisane, Selaelo Mametja, Mthokozisi Khoza, Mncengeli Sibanda, Motale Madigoe and Stanley Moloabi

Diabetes shared value model 121
Jane Ball, Claudia Pastellides, Nikash Bagirathi and Xolani Masombuka

STREAM IV: DIGITAL HEALTH AND DATA-DRIVEN PREVENTION 122

Data for value: integrating clinical and social risk data for population health management 122
Shannon Knight, Brian Ruff, Riedwaan Jabaar and Prince Chikezie

Improved sleep behaviours and healthcare costs amongst members of the Discovery Health Medical Scheme: a longitudinal observational study 137
Katherine Tregoning, Simon Henrie, Deepak Patel, Cassi Rubin, Lara Wayburne, Claudia Pastellides, Roseanne Harris and Mosima Mabunda

Association between recorded physical activity and cancer progression or mortality in individuals diagnosed with cancer in South Africa 150
Ntokozo Mabena, Nivash Rugbeer, Sandra Lehmann, Georgia Torres, Deepak Patel, Mosima Mabunda, Mike Greyling, Jane S Thornton, Yun-Hee Choi, Saverio Stranges and Jon S Patricios

ACKNOWLEDGEMENT OF THE RESEARCH SYMPOSIUM 151

ACKNOWLEDGEMENT OF INDABA PARTNERS 152



STREAM I
VALUE-BASED CARE MODELS

ONE GP: THE CASE FOR PRIMARY CARE COORDINATION IN CHRONIC CARE

Phumla Litlhakanyane, Claudia Pastellides, Aminda Bronkhorst, Amalia Theologides and Lara Wayburne

How to cite this article: Litlhakanyane, P., Pastellides, C., Bronkhorst, A., Theologides, A., & Wayburne, L. (2026). One GP: The case for primary care coordination in chronic care. *Council for Medical Schemes Research Journal*, *1*(1), 8–25.

ABSTRACT

Background: Continuity of care in primary health systems, anchored in an ongoing relationship with a single, regular general practitioner (GP), is widely recognised as a cornerstone of high-quality care, particularly for members with chronic conditions. However, the benefits of continuity of care through a single GP in the South African private healthcare setting remain largely unquantified.

Methods: Using a retrospective, observational analysis of claims data, this study evaluated the impact of GP continuity on healthcare utilisation and outcomes for members of a large South African medical scheme living with chronic diseases. Continuity was defined by whether members consistently consulted a single GP or multiple GPs over a one-year period. To address differences in case mix and potential selection bias, we applied both comprehensive risk adjustment and causal inference methods to estimate the effect of GP continuity on healthcare costs and utilisation.

Results: Members who consulted a single GP on an ongoing basis had lower risk-adjusted healthcare costs than those consulting multiple GPs, together with lower specialist, diagnostic (radiology and pathology) and hospital admission rates.

Conclusions: These findings provide the first in-depth South African evidence that strengthening primary care coordination, through the single "Primary Care Provider" GP nomination programme run by Discovery Health Medical Scheme, is associated with more efficient care and may reduce unnecessary healthcare expenditure for chronic populations.

KEYWORDS

Continuity of care; primary health care; general practitioner; care coordination; chronic disease; risk adjustment; causal inference; medical schemes; South Africa

INTRODUCTION

Background

Continuity of care with a single general practitioner (GP) is widely recognised as essential to improving outcomes, particularly for members living with chronic conditions. International studies over recent decades provide robust evidence that members who consistently see one GP experience better health outcomes and use healthcare more efficiently. A systematic review published in the *British Journal of General Practice*

concluded, with moderate certainty, that stronger GP–patient continuity probably reduces all-cause mortality, hospital admissions and emergency department visits, and that even relatively small improvements in continuity yield measurable benefits (Engström et al., 2025). Members who routinely consult a single GP are also more likely to receive preventive care, reflecting the importance of relational continuity for effective member engagement and disease management (Saultz & Lochner, 2005). Conversely, fragmented primary care, in which members consult multiple providers without a single coordinating GP, has been associated with duplicative testing, inconsistent treatment, greater use of emergency departments and higher overall healthcare costs (Joo, 2023).

These risks are especially pronounced for members living with chronic diseases, for whom the absence of continuous clinical oversight can lead to suboptimal treatment pathways, delayed identification of complications and avoidable hospital admissions.

The problem in context

Despite the well-established benefits of GP continuity, many health systems struggle to ensure that members have access to a regular primary care provider. In the United Kingdom, only around half of patients are reported to be able to consult the same GP consistently, reflecting a broader international decline in continuity of care (King's College London, 2022). In South Africa's private healthcare sector, primary care coordination has historically been limited. The Competition Commission South Africa's Health Market Inquiry (2019) identified fragmented provider markets and a lack of standardised quality measures, and found that members frequently move between providers without sustained continuity of care.

Discovery Health Medical Scheme (DHMS), South Africa's largest open medical scheme, faces the same challenge. The scheme covers more than 2.7 million members and accounts for just under 60% of the open medical scheme market. In 2025, approximately 24% of chronic DHMS lives consulted multiple GPs, and a further 15% of chronic members did not see a GP at all. Over the past 15 years, the scheme's membership has also aged significantly: the proportion of members over 60 years of age almost doubled, from 10% in 2010 to 19% in 2025, an average growth of roughly 4% a year. Older members typically have a high prevalence of chronic disease and multimorbidity, and their healthcare needs are rising accordingly. This demographic shift makes longitudinal GP oversight more important, yet many members continue to seek care episodically from different providers.

To strengthen primary care coordination for chronic members, DHMS introduced the Primary Care Provider (PCP) initiative in 2023. The initiative encourages each chronic member to nominate a single GP as their primary care coordinator, reinforced by incentives in the benefit design. Its explicit aim is to improve care coordination and outcomes through sustained patient–GP relationships.

Objectives and contribution of this study

Given the importance of continuity of care and the scarcity of South African evidence on the subject, this study investigates whether having one regular GP, rather than multiple GPs, is associated with better outcomes and lower costs for medical scheme members with chronic disease. Using retrospective demographic and healthcare claims data, we quantify the associations between GP continuity and four outcomes over a 12-month period: risk-adjusted healthcare expenditure, hospital admission rates, specialist and diagnostic referral rates, and follow-up care. We adjust for each member's age, sex, health plan and chronic disease profile, and apply causal inference methods to mitigate selection bias and strengthen causal interpretation of the results. By providing real-world evidence from a large South African medical scheme, the study addresses a gap in the continuity-of-care literature and offers policymakers and medical schemes insight into the value of GP-led care coordination in the South African private healthcare context.

LITERATURE REVIEW

Evidence from the World Health Organization (2018) indicates that continuity of care, measured by consistent visits with the same primary care doctor, is associated with improved outcomes, including fewer hospital admissions and emergency department visits. Systematic reviews indicate that fragmented care contributes to duplicative testing, inconsistent treatment, and increased healthcare costs (Joo, 2023). Similarly, Müskens et al. (2022) document widespread overuse of diagnostic testing in healthcare. Empirical studies demonstrate that higher continuity is associated with reduced mortality and lower hospitalisation rates (Engström et al., 2025). This is often measured using indices such as the Usual Provider of Care (UPC) index, a widely used measure of continuity defined as the proportion of visits to a single provider (Jee & Cabana, 2006).

In the South African context, Padayachee and Kathard (2026) provide an exploratory analysis of four national health policies, highlighting the extent to which continuity of care for persons with disabilities is embedded within policy frameworks. They found that continuity of care is acknowledged in policy but lacks clarity and implementation detail. The Health Market Inquiry (HMI) Final Findings and Recommendations Report (Competition Commission South Africa, 2019) identified significant fragmentation of care in South Africa's private healthcare system, which undermines care coordination and member outcomes. As reported in the CMS Industry Report (Council for Medical Schemes, 2024), GP services account for a relatively small component of medical scheme expenditure, despite modest growth in utilisation and spending. In contrast, hospital and specialist services account for the largest share of benefits (approximately 36% and 28%, respectively), indicating a structural reliance on higher-cost care. This pattern supports the CMS's broader policy position that strengthening primary care and GP-led gatekeeping is essential to improve efficiency and contain costs within the medical schemes environment.

Care coordination models such as Accountable Care Organisations (ACOs) provide additional evidence that structured coordination can improve outcomes, particularly for high-risk populations (NORC at the University of Chicago, 2024). However, much of the evidence in the literature is observational, which limits causal inference due to potential selection bias and unmeasured confounding. In addition, there is limited robust evidence on the effectiveness and the value of care coordination in the South African context. Consistent with international evidence, our findings indicate that establishing and maintaining an ongoing relationship with a single GP is associated with lower healthcare costs and more efficient use of care across the healthcare continuum.

This study contributes to the literature by:

- Providing real-world evidence from a large South African medical scheme;
- Examining differences in effects across heterogeneous member groups with appropriate risk adjustment; and
- Applying causal modelling techniques to strengthen the quantification of the association between consulting a single GP, compared with multiple GPs, and healthcare costs and utilisation.

The case for care coordination

In the absence of a clinician providing end-to-end oversight of the member's clinical history, healthcare delivery may become fragmented. This fragmentation may result in repeated investigations, suboptimal treatment, unclear accountability for follow-up, and delayed detection of clinical deterioration or complications.

Analysis of Discovery Health Medical Scheme (DHMS) data shows that members consulting multiple GPs incur higher overall healthcare costs than those managed by a single primary care provider, even after risk adjustment for age and clinical complexity. This cohort also demonstrates higher risk-adjusted hospital admission rates as well as increased utilisation of pathology and radiology services. The observed excess utilisation is assumed to have arisen from inefficiencies associated with fragmented care, including duplicated services, suboptimal care coordination, and delays in identifying and managing clinical risk. The risks associated with fragmented care increase with age and clinical complexity.

While episodic care may have limited short-term consequences for younger, healthier members, older members are more vulnerable to adverse outcomes due to multimorbidity and more intensive care needs. Over the past 15 years, the demographic profile of Discovery Health Medical Scheme (DHMS) has shifted substantially. The proportion of children and young adults has declined, while the proportion of members over the age of 60 years has almost doubled. This reflects a structural change in healthcare demand towards an older, more clinically complex population. Older members are significantly more likely to present with one or more chronic conditions and therefore experience higher hospital admission rates, increased reliance on chronic medication, and greater use of diagnostic services. Healthcare expenditure rises sharply with age, particularly beyond age 65 years. Members aged 75 years and older incur several multiples of the average per-life monthly cost of younger adults, driven predominantly by the ongoing management of chronic and multimorbid conditions. As member complexity increases, continuity of care with a single primary care provider becomes increasingly important to ensure appropriate clinical oversight, informed decision-making, and timely escalation of care when required. Members might consult with multiple healthcare professionals, in particular GPs, based on convenience, accessibility or perceived urgency rather than clinical need, which may undermine effective care coordination. In the absence of a structured care coordination framework, such patterns of care-seeking behaviour are likely to persist, contributing to suboptimal outcomes and inefficiencies.

To address this, DHMS implemented the Primary Care Provider (PCP) initiative in 2023, with the objective of strengthening care coordination by encouraging continuity of care through a designated GP as the primary care provider.

Enabling care coordination through Discovery Health's Primary Care Provider (PCP) initiative

Discovery Health's Primary Care Provider (PCP) initiative is designed to promote continuity of care by facilitating sustained relationships between members and a designated primary care provider. This is achieved through a combination of benefit design features and behavioural incentives, together with supportive digital tools, aimed at supporting appropriate healthcare utilisation and coordinated care delivery. Continuity with a single GP, who is further empowered through visibility of the member's healthcare utilisation and clinical history, enables more consistent clinical oversight and facilitates a more comprehensive understanding of the member's medical history, treatment patterns, and evolving care needs. This supports more coordinated decision-making and reduces the likelihood of duplicated investigations, inconsistent clinical advice, and inefficient referral pathways. The PCP initiative is operationalised through a formal nomination process, whereby members elect a designated primary care provider responsible for overseeing and coordinating their ongoing care. This mechanism establishes a clear point of clinical accountability, while preserving member choice within defined scheme rules.

Members can nominate a PCP through several channels: the scheme's digital application, telephonically through a service agent, or through the member's participating GP. All scheme members are encouraged to select and nominate a PCP. Awareness is driven through broad awareness campaigns as well as through targeted communication at the time that GP services are used. To reinforce these care coordination objectives, benefit design incorporates a copayment when chronic members consult a GP other than their nominated PCP. Exceptions apply in defined circumstances, such as emergency care or when defined access

constraints are present. This approach aligns financial incentives with appropriate utilisation, supports continuity of care and care coordination, and aims to reduce avoidable fragmentation, while maintaining reasonable access to care in line with regulatory requirements. From a behavioural perspective, the use of copayments to encourage member alignment with a nominated PCP is supported by a substantial body of literature on health-related incentives and decision-making. Behavioural economics suggests that individuals respond predictably to financial signals, with even relatively small cost differentials influencing healthcare utilisation choices (Thaler & Sunstein, 2008). Other behavioural levers adopted to support this initiative include ongoing reminders to members to use their PCP as the entry point to the healthcare system and facilitating easy access to the member's PCP (for example, by offering online appointment booking and making contact details readily available). The PCP initiative is intended to support appropriate navigation of the healthcare system, facilitate timely access to primary care, and enable more effective management of care across settings, with the objective of improving clinical outcomes and the overall care experience for members.

METHODOLOGY

Investigating the impact of care coordination for DHMS chronic members

Data description and population characteristics

The study population comprises medical scheme members with registered chronic conditions defined according to the Chronic Disease List of the Prescribed Minimum Benefits, observed over the period from 1 January 2023 to 31 December 2025. The research followed a three-period design from 2023 to 2025, with the following time horizons:

- Baseline period (2023): measurement of demographic characteristics, baseline healthcare utilisation and initial GP-seeking behaviour;
- Intervention period (2024): observed changes in GP allocation strength and care coordination behaviour; and
- Outcome period (2025): assessment of subsequent healthcare utilisation and costs after the intervention.

This structure allows for evaluation of the association between changes in GP care coordination and subsequent healthcare utilisation and costs while accounting for baseline differences. The final analytical sample included 452,793 members who met all inclusion criteria for the causal modelling analysis and had continuous observation across the three-year study period. Members were included in the analysis only if they had recorded claims data and at least one general practitioner (GP) visit in each year of the three-year study period, ensuring continuous observation and a consistent basis for evaluating care coordination patterns over time.

Members were also required to meet all of the following criteria:

- Registered on the Chronic Illness Benefit (CIB)
- Nominated a Primary Care Provider (PCP)
- Continuous membership from 2023 to 2025

Members were excluded if they had:

- Pregnancy-related claims
- Trauma-related events
- An oncology registration

These exclusions were applied to ensure that the analysis isolates the effects of care coordination rather than reflecting healthcare utilisation driven by acute, non-discretionary or episodic events.

Data preparation

Demographic and healthcare claims data for the observation period were used in this analysis. Members with no general practitioner (GP) utilisation were excluded from the analysis to ensure a valid comparison of care coordination patterns between members who saw one GP and those who saw multiple GPs. Categorical variables were transformed into numerical representations for the modelling framework. To limit the influence of extreme values and reduce the distortion introduced by outliers, particularly those driven by rare, high-cost events, cost variables were capped, with total healthcare costs truncated at R500,000 and in-hospital costs capped at R200,000. This improves the stability and robustness of the estimates. Negative values, arising from adjustments or data anomalies, were truncated to zero. Finally, duplicate records were removed to ensure a one-to-one mapping of members, using the unique identifier across all observations.

Care coordination patterns over time

Care coordination patterns were assessed using GP allocation strength, defined as consulting a single GP, consulting multiple GPs, or having no GP allocation. This categorisation is shown in Table 1 below. Among members who were consulting a single GP in 2023, the majority (88%) continued to consult a single GP in 2025, indicating strong continuity of care. However, smaller proportions transitioned to consulting multiple GPs (10%) or had no GP allocation (2%). In contrast, members who were consulting multiple GPs in 2023 exhibited more variation in care patterns by 2025. Approximately 46% transitioned to consulting a single GP, while 48% continued consulting multiple GPs and 6% had no GP allocation. This group represents the population experiencing the biggest improvements in care coordination. Members with no GP allocation in 2023 experienced a more dispersed healthcare utilisation pattern in 2025, with 50% transitioning to consulting a single GP, while 25% moved to consulting multiple GPs and 25% remained without a GP allocation.

Table 1: Care coordination trends from 2023 to 2025

	Status in 2025		
Care coordination status in 2023	Seeing one GP	Seeing many GPs	No allocation
Seeing one GP	264,406 (88%)	29,687 (10%)	6,361 (2%)
Seeing many GPs	48,662 (46%)	50,798 (48%)	6,380 (6%)
No allocation	23,121 (50%)	11,563 (25%)	11,815 (25%)

Note: Values are numbers of members, with row percentages in parentheses.

Demographic profile of the study population

The demographic profile of the study population shown in Table 2 below indicates a predominantly older and clinically complex cohort. The population is skewed towards older age groups:

- Members aged 60 years and older account for approximately 45% of the total population
- Members aged 40–60 years represent approximately 38% of the total population
- Younger members (<40 years) comprise a smaller proportion (17%) of the sample

Across care coordination groups, members consulting a single GP tend to be slightly older, with 46% aged 60+, compared to 41% among those consulting multiple GPs and 38% among those with no GP allocation. The gender distribution is broadly balanced, with a slight majority of female members (54%) in the total

population. This pattern is consistent across care coordination categories, with only modest variation. Members are distributed across a range of benefit options, with the largest concentrations observed in:

- Saver plans (approximately 51%)
- KeyCare plans (approximately 14%)

Members consulting multiple GPs show slightly higher representation in Comprehensive and Priority plans. This likely reflects differences in benefit design, as these plans offer greater flexibility in provider choice, allowing members to consult multiple GPs without strict referral requirements. In contrast, lower-tier plans are typically subject to referral pathways and network restrictions, which limit access to designated providers and encourage more coordinated care through a single GP.

Clinical profile of the study population

The study population consists of:

- Approximately 55% of members have one or more Chronic Disease List (CDL) conditions.
- The remaining 45% have no CDL condition.

Members consulting a single GP exhibit a higher prevalence of chronic disease (58% CDL chronic) compared to those consulting multiple GPs (54%) and those with no allocation (33%), suggesting that members with more complex clinical needs may be more likely to engage in coordinated care with a single provider.

Table 2: Demographic profile of the study population

Characteristic	Category	Care coordination status (2025)			
		Seeing one GP	Seeing many GPs	No allocation	Total DHMS
Age (years)	<30	6%	10%	11%	7%
	30–40	9%	12%	12%	10%
	40–60	39%	37%	40%	38%
	60+	46%	41%	38%	45%
Gender	Female	54%	56%	49%	54%
	Male	46%	44%	51%	46%
Plan	Executive	1%	1%	1%	1%
	Comprehensive	8%	15%	8%	10%
	Priority	7%	11%	6%	8%
	Saver	52%	57%	53%	51%
	Core	10%	7%	23%	10%
	Smart	7%	8%	7%	7%
	KeyCare	15%	1%	3%	14%
Chronic status	CDL chronic	58%	54%	33%	55%
	CDL non-chronic	42%	46%	67%	45%

Note: Values are column percentages. CDL = Chronic Disease List; DHMS = Discovery Health Medical Scheme.

DEFINITION OF CONTINUITY OF CARE

Continuity of care was assessed using an attribution approach analogous to the Accountable Care Organisation (ACO) model as defined by the Centers for Medicare & Medicaid Services (CMS, n.d.), whereby members are assigned to providers based on utilisation. Specifically, an allocation method was applied based on patterns of out-of-hospital (OH) general practitioner (GP) utilisation over a 12-month rolling period, with each member assigned to a primary GP defined as the provider most frequently consulted during this period. Continuity was categorised according to the concentration of visits with a single GP as follows:

- Single GP: All GP visits are with a single GP.
- Multiple GPs: Members who consulted more than one GP, further classified as:
  - Minority concentration: Fewer than 60% of visits are with a single GP.
  - Majority concentration: 60% or more of visits are with a single GP.

Members may consult multiple GPs for a variety of practical and clinical reasons, including seeking care outside of their usual GP's consulting hours, accessing urgent or after-hours services through casualty or walk-in providers, or visiting different providers due to availability, convenience, or geographic accessibility.

Risk adjustment

Outcomes were assessed by comparing risk-adjusted claims costs per life per month (PLPM) and healthcare utilisation patterns between members who consulted a single GP and those who consulted multiple GPs during the study period. Utilisation measures included GP and specialist visit rates, hospital admissions, referrals for pathology and radiology, and medicine adherence.

To make meaningful comparisons of the health outcomes between members with different health profiles, we adjusted for confounding factors that could influence a member's risk of adverse outcomes such as:

- A member's age and sex;
- A member's medical aid plan; and
- Whether a member has at least one Chronic Disease List (CDL) condition.

Chronic disease burden is captured through 27 Chronic Disease List (CDL) condition indicators, and the structure of the risk adjustment model is detailed in Annexure A.

Measuring the association of improved care coordination using a causal framework

While risk adjustment helps control for observable pre-treatment differences (e.g., age, sex, plan, and chronic disease burden), it remains limited to correlational comparisons and may not fully address selection bias. To strengthen the analysis, we applied causal inference techniques to more formally account for differences between members receiving varying levels of care coordination. These methods enable estimation of the expected differences in healthcare utilisation and costs under alternative care patterns, while reducing bias arising from systematic differences in member characteristics. As such, the approach moves beyond simple adjusted comparisons and provides a more robust assessment of the association between improved care coordination and healthcare outcomes.

The following section outlines the specific causal modelling approach applied in this analysis.

A retrospective study design was implemented using causal inference techniques to estimate the association between improved care coordination with a primary care provider and subsequent healthcare costs at a

population level.

The analysis adopts a three-year longitudinal data structure:

- **Baseline (Evaluation Period): 1 January 2023 – 31 December 2023.** Baseline data were collected on members' "prior" risk profiles, historical utilisation and care coordination patterns, primarily among members seeing multiple GPs.
- **Intervention Period: 1 January 2024 – 31 December 2024.** Period used to observe changes in GP allocation strength and care coordination behaviour.
- **Outcome Period: 1 January 2025 – 31 December 2025.** Data were collected to measure subsequent healthcare costs and utilisation outcomes after the intervention.

All causal modelling analyses were conducted in R, using specialised packages for causal inference and machine learning.

CAUSAL MODEL STRUCTURE

The association between improved care coordination during 2024 (t_1) and total healthcare costs in 2025 (t_2) is estimated by adjusting for the observed confounders measured in 2023 (t_0), as shown in Figure 1.

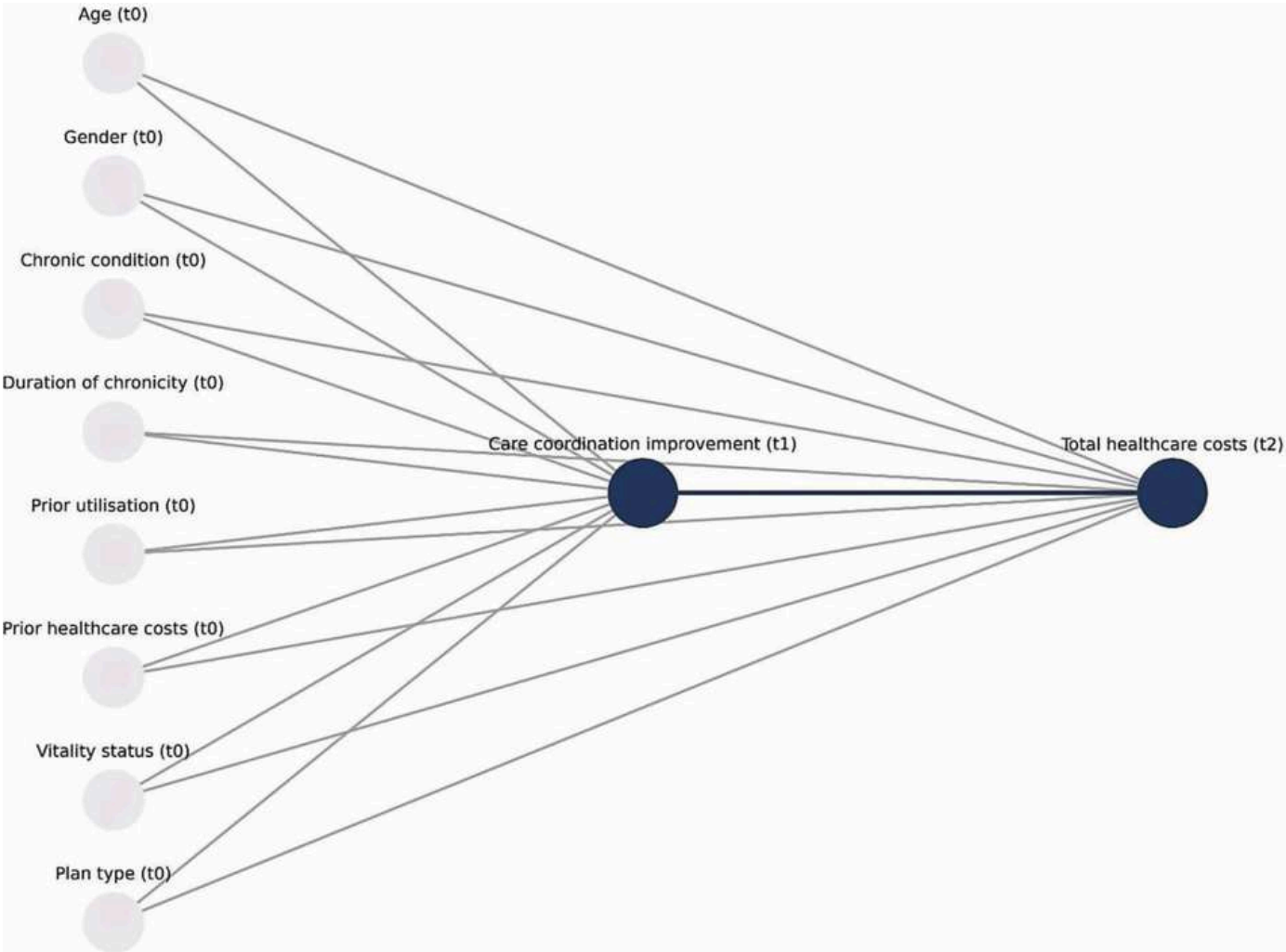

Figure 1: Directed acyclic graph illustrating the causal framework for care coordination

MODEL SPECIFICATION

The treatment group consists of members transitioning from low continuity of care, defined as consulting multiple general practitioners (GPs) in 2023, to high continuity of care, defined as consulting a single GP in 2025. In contrast, the control group includes members who continue to experience fragmented care by consulting multiple GPs or who do not engage with GP services over the study period.

To estimate the effect of improved care coordination on healthcare costs, a causal inference framework based on a Generalised Random Forest was applied. This approach utilises an ensemble of decision trees (n = 1,000) to model complex, non-linear relationships within the data. The causal forest algorithm is specifically designed to estimate heterogeneous treatment effects, allowing the impact of improved GP continuity to vary across individuals while adjusting for high-dimensional confounding factors.

The causal model estimates the Average Treatment Effect (ATE) as the average reduction in total costs associated with improved GP coordination (a positive ATE therefore denotes a saving), defined as ATE = E[Y(0) – Y(1)].

RESULTS

The relationship between care coordination and healthcare utilisation and cost outcomes

Table 3 presents a consolidated summary of healthcare utilisation and cost outcomes by level of GP continuity, with all measures risk-adjusted for demographic characteristics and clinical complexity. These ratios are derived deterministically from risk-adjusted values and therefore have no associated standard errors; statistical inference is provided by the causal model, for which confidence intervals and p-values are reported (Table 4).

Table 3: Cost and utilisation multiples for members seeing one GP compared with members seeing multiple GPs (reference = 1)

Measure	Seeing many GPs (reference)	Seeing one GP
Risk-adjusted cost per life per month	1.00	0.83
Risk-adjusted specialist visit rates	1.00	0.75
Risk-adjusted referred pathology rates	1.00	0.94
Risk-adjusted radiology referral rates	1.00	0.77
Risk-adjusted 30-day GP follow-up visit rates following hospital admission	1.00	1.10
Risk-adjusted admission rates	1.00	0.84
Risk-adjusted 30-day readmission rates	1.00	0.92

Overall, members who consult a single GP consistently demonstrate more efficient healthcare utilisation patterns and lower costs compared to those consulting multiple GPs. Most notably, total healthcare costs per life per month (PLPM) are approximately 17% lower among members seeing a single GP, reflecting a material reduction in overall healthcare expenditure associated with improved care coordination.

These cost differences are accompanied by consistent reductions across multiple benefit categories. Members consulting with a single GP incur between 20% and 30% lower PLPM costs for medication, consultations, pathology, clinical procedures, and radiology services compared to members consulting

multiple GPs. This pattern is consistent with reduced utilisation across multiple care settings (in-hospital and out-of-hospital).

Members consulting a single GP exhibit 25% lower specialist visit rates, indicating reduced reliance on secondary care services. This suggests that continuity of care supports more comprehensive management within primary care, potentially reducing the need for specialist referral.

Similarly, diagnostic utilisation is lower among members with higher care coordination. Pathology referral rates are approximately 6% lower, while radiology utilisation is reduced by approximately 23% among members consulting a single GP. These patterns are consistent with more coordinated and informed clinical decision-making, where access to an integrated member history reduces the likelihood of duplicated or unnecessary diagnostic testing.

Improvements in care coordination are also reflected in post-acute care management. Members consulting a single GP are approximately 10% more likely to attend a follow-up consultation within 30 days of hospital discharge, indicating better continuity between inpatient and outpatient care. This enhanced follow-up may contribute to improved clinical monitoring and early identification of complications.

Consistent with this, hospital utilisation outcomes are also more favourable among members with higher continuity of care. Risk-adjusted hospital admission rates are approximately 16% lower among members consulting a single GP, suggesting that coordinated primary care may support earlier intervention and ongoing disease management.

Readmissions are defined as unplanned returns to acute care within 30 days of hospital discharge. The 30-day readmission rate is approximately 8% lower, indicating more effective post-discharge care and reduced likelihood of avoidable hospital returns.

Taken together, these findings indicate that improved GP coordination is associated with broad-based reductions in healthcare utilisation and costs, spanning primary care, specialist services, diagnostics, and hospital-based care. The consistency of these effects across multiple metrics supports the interpretation that continuity of care is associated with more efficient and better-coordinated healthcare delivery.

A key limitation in observational analyses of care coordination is the potential for selection bias, whereby differences in outcomes may reflect underlying differences in member characteristics rather than the effect of care coordination itself. For example, healthier members may be more likely to establish and maintain sustained relationships with a single GP, potentially leading to an overestimation of the benefits associated with care coordination.

To address this, we applied comprehensive risk adjustment using pre-treatment covariates, including age, sex, plan and chronic disease burden, to account for observable differences in baseline risk profiles.

Figures 2 and 3 show that admission rate differences between single-GP and multiple-GP users increase with clinical complexity and age. The relative gap rises from 9% for members with one chronic condition to 25% for those with four or more chronic conditions, with a similar pattern across age groups above 60 years.

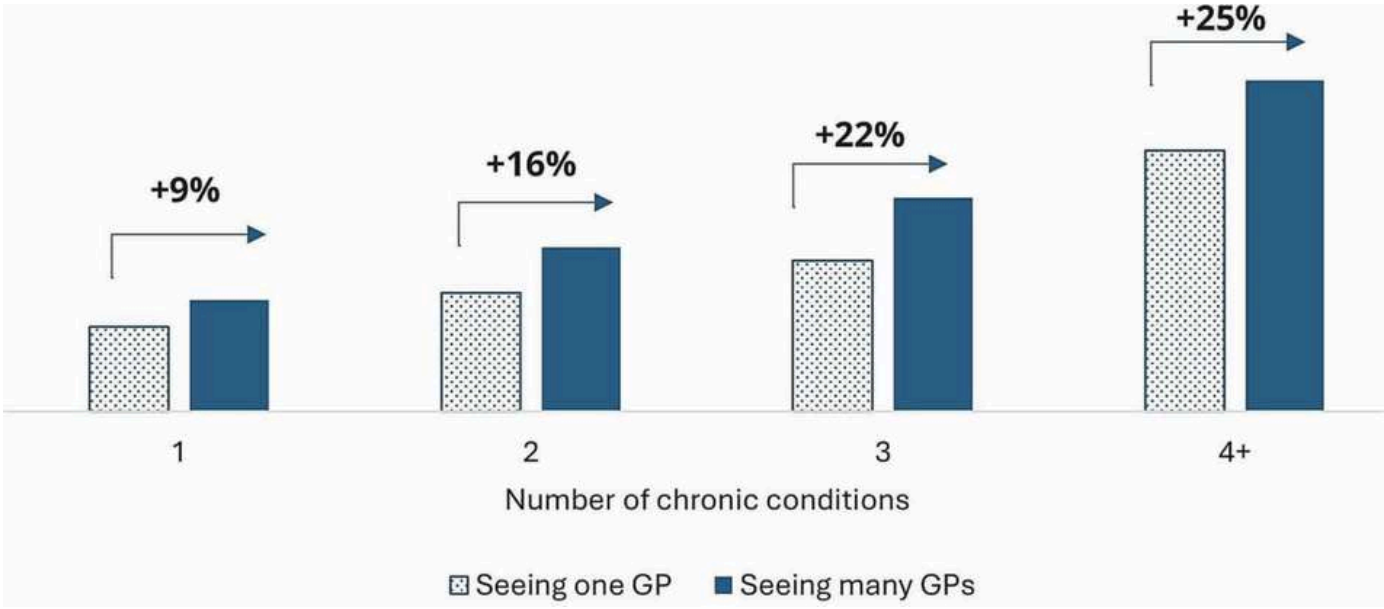

Figure 2: Risk-adjusted admission rates by number of chronic conditions

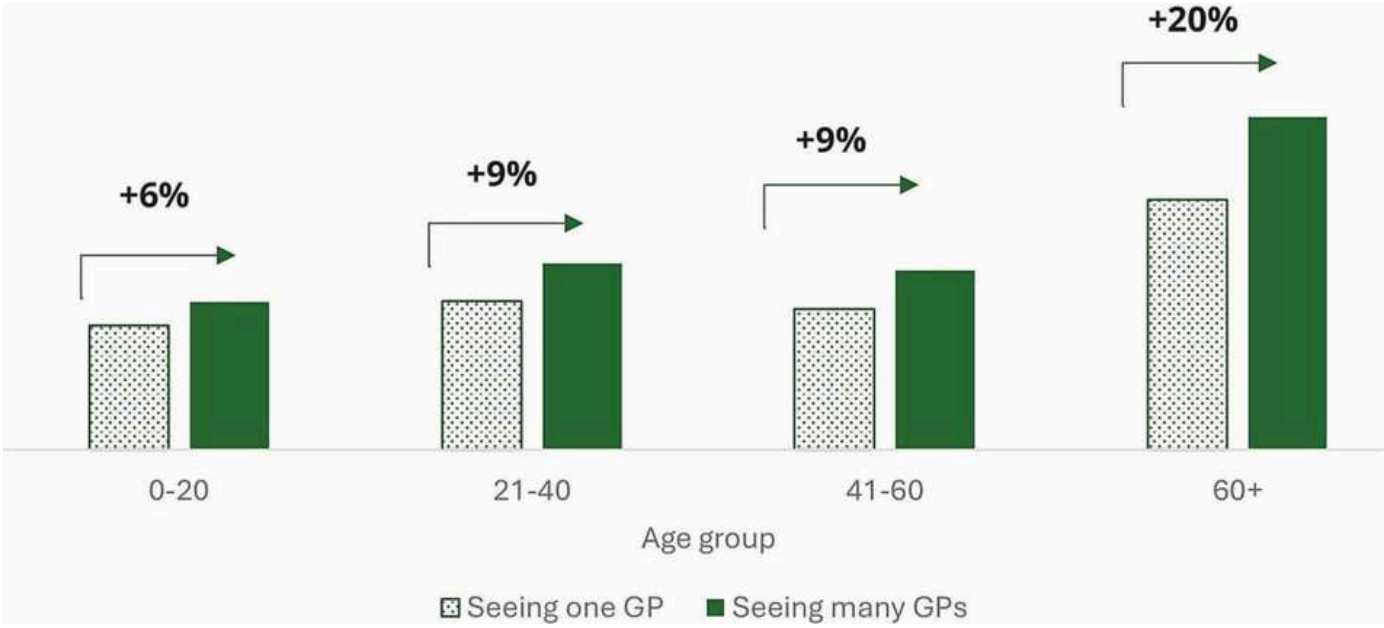

Figure 3: Risk-adjusted admission rates by age group

These results indicate more pronounced effects among higher-risk, clinically complex populations, for whom selection into coordinated care is less likely to reflect inherently favourable underlying health status, thereby strengthening the inference that observed benefits are attributable to care coordination itself.

This finding is consistent with evidence from the evaluation of the Next Generation Accountable Care Organisation (NGACO) model (NORC at the University of Chicago, 2024), which reported that care coordination for high-risk members with multiple chronic conditions was associated with cost savings and reductions in all-cause inpatient admissions.

Estimated average treatment effect from causal inference

The estimated average treatment effect (ATE) reflects the average change in total healthcare costs associated with improvements in GP coordination, measured as members transitioning from seeing multiple GPs to seeing only a single, consistent GP.

The results indicate that members who transitioned from seeing multiple GPs to seeing one GP experienced a meaningful reduction in healthcare costs. On average, improved continuity of care was associated with an approximate 2.7% reduction in total healthcare costs on a per life per month (PLPM) basis (Table 4).

Table 4: Average treatment effect of moving to a single GP on healthcare costs (PLPM)

Cohort	ATE (R)	PLPM savings (R)	% Efficiency (savings/total costs)	95% confidence interval	p-value
Minority → One GP	R1 863	R270	5.30%	[R972; R2 754]	< 0.0001
Majority → One GP	R440	R60	1.20%	[–R307; R1 187]	0.25

Further analysis showed significant heterogeneity in treatment effects across different member groups. Members with highly fragmented baseline care, defined as those seeing many GPs with no dominant GP in the baseline period, experienced the largest cost reductions in the observation period. For this group of members, transitioning to a single GP was associated with cost savings of 5.3% of total healthcare costs. An average treatment effect (ATE) of R1 863 was observed with the following 95% confidence interval: [R972; R2 754]. The associated p-value is less than 0.0001, indicating strong evidence against the null hypothesis of no effect of the intervention on the outcome for members seeing many GPs, none of whom accounted for a majority of visits.

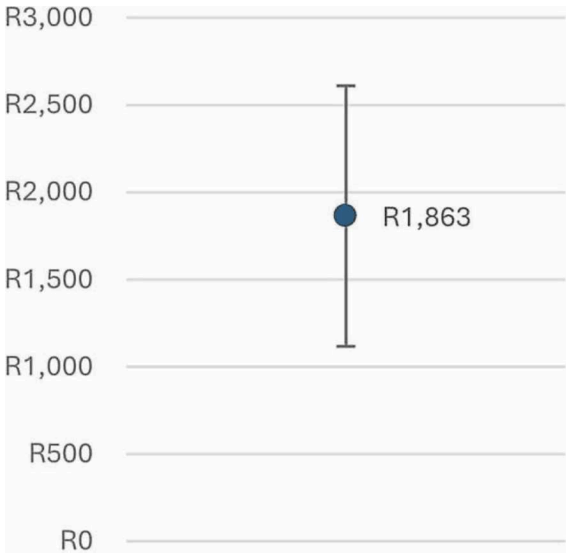

Figure 4: Average treatment effect of seeing only one GP rather than multiple GPs (none with majority)

In contrast, members who already showed some degree of continuity at baseline, i.e., with a dominant GP accounting for most of their visits, had a smaller estimated saving that was not statistically significant. For this group, transitioning from seeing many GPs (one with a majority of visits) to seeing only one GP was associated with an approximate 1.2% reduction in total healthcare costs. An ATE of R440 was observed with the following 95% confidence interval: [–R307; R1 187]. The associated p-value is approximately 0.25, indicating that the effect of the intervention on the outcome is not statistically significant for those seeing many GPs (one with a majority of visits).

The absence of statistical significance does not demonstrate that there are no savings where care is already largely coordinated; the study may have been underpowered to detect smaller savings in this group.

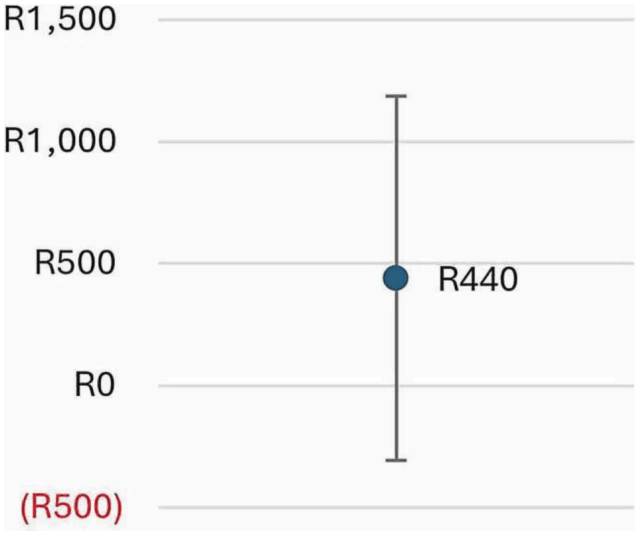

Figure 5: Average treatment effect of seeing only one GP rather than multiple GPs (one with majority)

Overall, these results indicate that improvements in care coordination are associated with meaningful reductions in healthcare costs. However, the magnitude of the savings depends on the extent of fragmentation in baseline care patterns. The intervention appears to be most effective for the group with the highest level of fragmentation, where it yields a statistically significant positive effect. In contrast, there is no reliable evidence of an effect for the group that already tends to concentrate most of their visits with a single GP, despite still consulting multiple providers.

DISCUSSION

This study provides empirical evidence that improved continuity of care, characterised by consistent engagement with a single general practitioner (GP), is associated with lower healthcare costs and more efficient utilisation within the South African private healthcare setting. Across both risk-adjusted descriptive analyses and causal modelling approaches, a clear and consistent pattern emerges: members receiving more coordinated primary care incur lower overall healthcare costs, demonstrate reduced utilisation of higher levels of care (including specialist services and hospitalisation) and exhibit improved continuity across the care pathway.

The risk-adjusted findings indicate broad-based efficiency gains associated with GP continuity. Members consulting a single GP consistently demonstrate lower healthcare expenditure, fewer specialist consultations, reduced diagnostic utilisation, and lower rates of hospital admission and readmission. Importantly, these effects are not confined to a single outcome but are observed across multiple dimensions of healthcare utilisation, suggesting that continuity contributes to systemic efficiency rather than simply redistributing utilisation across service categories. This is strongly aligned with international evidence, which links continuity of care to lower hospital utilisation, improved chronic disease management, and more efficient allocation of healthcare resources. The finding that the largest effects are concentrated among higher-risk, more clinically complex members further corroborates evidence from coordinated care models, such as Accountable Care Organisations, where care coordination yields the greatest benefits for populations with multiple chronic conditions.

The causal inference analysis further strengthens these findings by more comprehensively accounting for selection bias and examining the relationship between improvements in care coordination and subsequent outcomes. The results indicate that a transition from fragmented care (seeing multiple GPs) to more coordinated care (seeing a single GP) is associated with a measurable reduction in healthcare costs, with an average decrease of approximately 2.7% in total healthcare costs per life per month. Notably, the magnitude of this association is heterogeneous: members with highly fragmented baseline care experience the largest

reductions in costs, while those with partial continuity show a smaller estimated saving that was not statistically significant. This pattern suggests that the extent of improvement in care coordination is closely linked to the magnitude of observed efficiency gains.

From a policy and practice perspective, these findings have important implications for medical schemes and health systems. The demonstrated association between GP continuity and improved efficiency supports interventions that encourage sustained relationships between members and primary care providers.

More broadly, the results highlight the value of strengthening primary care systems and reinforcing the role of GPs in coordinating care, particularly for populations with high chronic disease burden. They also highlight the importance of creating structural frameworks that support and encourage the use of a primary care GP as the entry point to the healthcare system, creating sustainable changes in how healthcare services are accessed.

Such approaches have the potential to improve long-term health outcomes while enhancing the sustainability of healthcare financing. This study contributes to the literature by providing one of the first quantitative assessments of the impact of care coordination in the South African context, combining large-scale longitudinal data with advanced causal inference techniques to estimate the real-world effects of improved GP continuity.

LIMITATIONS

While this study provides evidence on the benefits of GP continuity, several limitations should be noted when interpreting the results.

First, the analysis is restricted to members of a single large medical scheme, which may limit generalisability to the broader South African population with different access constraints and socioeconomic contexts. Second, outcomes are evaluated over a one-year period following changes in care coordination, which may understate the longer-term benefits of sustained GP continuity. Third, the measure of continuity does not account for members who change their primary GP during the year, potentially leading to conservative estimates of its benefits. Fourth, despite comprehensive risk adjustment and the use of causal inference methods, the analysis remains observational and may be subject to residual confounding from unobserved factors such as health-seeking behaviour, member preferences, and provider quality. Alternative explanations for the lower utilisation among members who consult a single GP should also be considered, including behavioural differences between members, variation in provider practice, and the effects of access constraints or plan design.

Finally, the study focuses on cost and utilisation outcomes and does not directly measure clinical outcomes or patient-reported experiences, which may provide a more complete assessment of the impact of care coordination. Despite these limitations, the consistency of findings across multiple outcomes, the use of a large longitudinal dataset, and the application of causal inference techniques support the overall conclusions.

CONCLUSION

This analysis integrates evidence from real-world Discovery Health Medical Scheme member data and the international literature. The results show that continuity of care delivered through a patient-centred single-GP model is associated with more efficient healthcare utilisation and lower costs. This single-GP framework enables more efficient use of healthcare resources and may support better long-term health outcomes.

Among members living with chronic conditions, consistent engagement with a single GP is associated with the following outcomes: lower risk-adjusted costs per life per month as a result of fewer specialist and

diagnostic referrals, higher rates of post-hospitalisation follow-up, and lower admission and readmission rates. These associations are most pronounced among members with greater clinical complexity and multiple chronic conditions, indicating that continuity of care confers increasing benefits as health needs increase and intensify. From a healthcare system perspective, coordinated primary care supports more effective clinical oversight, improved care navigation, and earlier identification and management of emerging health risks.

Discovery Health's Primary Care Provider (PCP) initiative is structured to advance these objectives by strengthening continuity of care while maintaining member choice within defined scheme rules. As healthcare demand continues to shift towards the long-term management of chronic and multimorbid conditions in an ageing population, coordinated primary care anchored in a single GP relationship represents a lever for improving care quality, enhancing utilisation efficiency, and supporting the sustainability of medical schemes.

ANNEXURE A: RISK ADJUSTMENT MODEL

Purpose of the model

This model estimates expected healthcare costs for members so that different groups can be compared fairly. It adjusts for differences in health risk (for example, age or chronic illness), allowing "like for like" comparisons between cohorts.

The output of the model is a cost weight:

- A value above 1 means higher than average expected cost
- A value below 1 means lower than average expected cost

How the data is structured

The model works at a member-by-month level. Each row represents one member in one month, along with information that explains their expected healthcare cost.

The key inputs include:

- Age: how old the member is
- Gender: the member's gender
- Chronic conditions (27 CDL indicators)
- Cancer-related information (where applicable):
  - Type of primary cancer
  - How long the member has been enrolled in the oncology programme
  - Whether the cancer has spread (metastatic vs non-metastatic)
- Plan code: the plan the member is on

These factors are used to capture differences in health status and benefit design that drive cost.

What the model predicts

The model predicts a cost weight rather than a raw money amount. This is calculated as:

Member's cost per life per month (PLPM) ÷ Average PLPM for that year

Because costs are expressed relative to the yearly average:

- Results can be compared across different years
- No adjustment for medical inflation is required

Model type

The model uses a Random Forest, which is a machine learning method that:

- Builds many decision trees
- Combines their results to produce stable and accurate predictions
- Handles complex, non-linear relationships between risk factors and cost

This makes it well suited to healthcare cost modelling, where costs are influenced by many interacting variables.

Training and validation

- Training data: Full year of 2017 data
- Testing and validation: 2016 data, split evenly into two halves

This approach ensures the model:

- Learns patterns from a complete and stable dataset
- Is validated on data from a different year to check performance

REFERENCE LIST

Centers for Medicare & Medicaid Services. (n.d.). *Accountable care and accountable care organizations*. Retrieved June 10, 2026, from https://www.cms.gov/priorities-innovation-key-concepts-accountable-care-accountable-care-organizations

Competition Commission South Africa. (2019). *Health Market Inquiry: Final findings and recommendations report*. https://www.compcom.co.za/wp-content/uploads/2020/01/Final-Findings-and-recommendations-report-Health-Market-Inquiry.pdf

Council for Medical Schemes. (2024). *The medical schemes industry in 2024* (Industry Report 2024).

Engström, S. G., André, M., Arvidsson, E., Östgren, C. J., Troein, M., & Borgquist, L. (2025). Personal GP continuity improves healthcare outcomes in primary care populations: A systematic review. *British Journal of General Practice*, *75*(757), e518–e525. https://doi.org/10.3399/BJGP.2024.0568

Jee, S. H., & Cabana, M. D. (2006). Indices for continuity of care: A systematic review of the literature. *Medical Care Research and Review*, *63*(2), 158–188. https://doi.org/10.1177/1077558705285294

Joo, J. Y. (2023). Fragmented care and chronic illness patient outcomes: A systematic review. *Nursing Open*, *10*(6), 3460–3473. https://doi.org/10.1002/nop2.1607

King's College London. (2022, July 7). *Only half of patients regularly see the same GP* [News release]. https://www.kcl.ac.uk/news/only-half-of-patients-regularly-see-the-same-gp

Müskens, J. L. J. M., Kool, R. B., van Dulmen, S. A., & Westert, G. P. (2022). Overuse of diagnostic testing in healthcare: A systematic review. *BMJ Quality & Safety*, *31*(1), 54–63. https://doi.org/10.1136/bmjqs-2020-012576

NORC at the University of Chicago. (2024). *Next Generation Accountable Care Organization Model evaluation: Sixth and final evaluation report*. Centers for Medicare & Medicaid Services. https://www.cms.gov/priorities/innovation/data-and-reports/2024/nextgenaco-sixthevalrpt

Padayachee, T., & Kathard, H. (2026). Continuity of care for persons with disabilities in South Africa: An exploratory analysis of four health policies. *Disabilities*, *6*(1), Article 19. https://doi.org/10.3390/disabilities6010019

Saultz, J. W., & Lochner, J. (2005). Interpersonal continuity of care and care outcomes: A critical review. *Annals of Family Medicine*, *3*(2), 159–166. https://doi.org/10.1370/afm.285

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

World Health Organization. (2018). *Continuity and coordination of care: A practice brief to support implementation of the WHO Framework on integrated people-centred health services*. https://www.who.int/publications/i/item/9789241514033

DECLARATIONS

Conflict of interest

The authors are employees of Discovery Health, the organisation which provided the data used in this study. While all analyses were conducted using rigorous, data-driven methods and findings are reported irrespective of commercial considerations, readers should be aware of this potential conflict of interest. We encourage independent replication of our results where feasible.

Acknowledgements

The authors wish to acknowledge the valuable support provided by colleagues at Discovery Health in the preparation and proofreading of this paper. The authors also acknowledge that they are employees of Discovery Health (Pty) Ltd, an accredited medical scheme administrator.

Author contributions

Phumla Litlhakanyane: Conceptualisation, Methodology, Formal Analysis; Writing. Claudia Pastellides: Conceptualisation, Methodology, Formal Analysis; Writing. Aminda Bronkhorst: Conceptualisation; Writing, Review & Editing. Amalia Theologides: Literature; Writing, Validation, Review & Editing. Lara Wayburne: Conceptualisation; Writing, Supervision, Review & Editing. All authors have read and approved the final version of this manuscript.

Data availability statement

The data supporting the findings of this study are available from the authors upon reasonable request, subject to ethical approval and institutional regulations.

Declaration of competing interest

The authors are employees of Discovery Health (Pty) Ltd, the administrator of Discovery Health Medical Scheme, which provided the data and administers the Primary Care Provider initiative evaluated in this study (see Conflict of interest). No other competing interests are declared.

Declaration of ethics

The authors declare that the research and writing of this manuscript adhere to ethical standards of research and publication, in accordance with scientific principles, and are free from plagiarism.

Declaration of assistive technologies in the writing process

The authors declare that generative artificial intelligence (GenAI) and AI-assisted tools were used in a limited and responsible manner during the preparation of this manuscript. The authors take full responsibility for the final content of this manuscript.



STREAM I
VALUE-BASED CARE MODELS

IMPLEMENTING INTEGRATED PRIMARY CARE HUBS IN PRIVATE HEALTHCARE: LESSONS FROM A DECADE OF PRACTICE IN SOUTH AFRICA

Hellen Nkwagatse and Belinda Richards

How to cite this article: Nkwagatse, H., & Richards, B. (2026). Implementing integrated primary care hubs in private healthcare: Lessons from a decade of practice in South Africa. *Council for Medical Schemes Research Journal*, *1*(1), 26–38.

ABSTRACT

Background: The fragmentation of primary care in South Africa's private healthcare sector has resulted in hospital-centric care delivery, erosion of general practitioner scope and capabilities, and poor coordination for patients with complex needs. Three decades of benefit design favouring inpatient services have systematically undermined outpatient care, leaving patients unable to navigate an increasingly fragmented healthcare landscape. International evidence supports integrated, team-based primary care models, and the replacement of prevalent fee-for-service environments remains crucial.

Methods: This practice account describes how a healthcare services management organisation in South Africa iteratively developed an integrated care hub model over a decade. The model organises general practices into coach-led local multidisciplinary teams supported by centralised enabling functions, including analytics,, quality improvement and care coordination infrastructure. Beginning with pilot sites in 2016, the approach evolved through continuous learning into a national programme serving over 135,000 members of a large restricted medical scheme across five provinces.

Results: Three fundamental lessons emerged from a decade of implementation. First, horizontal integration among primary care colleagues (GPs with care coordinators, allied health professionals and local services such as mental health, sub-acute and hospice care) must precede vertical integration with specialists and hospitals; trust within hubs is the foundation for broader coordination. Value-based care (VBC) contracts create the conditions but cannot substitute for building teams. Second, transforming practice organisation and clinical practice requires sustained coaching; behaviours embedded over decades cannot be changed by workshops and incentives alone. Third, technology enables but cannot drive transformation; it supports the shared-purpose structures that drive change.

Conclusions: This experience demonstrates that integrated primary care hubs can be implemented within existing private healthcare structures when organisational infrastructure supports practice transformation. The approach offers lessons for health systems seeking to strengthen primary care as a foundation for universal health coverage (UHC).

KEYWORDS

Integrated care; primary health care; multidisciplinary teams; value-based care; practice transformation; South Africa

INTRODUCTION

The problem

Primary care in South Africa's private healthcare sector faces a crisis of fragmentation. Over three decades, benefit design has systematically favoured hospital-based care over outpatient services, creating perverse incentives that have reshaped clinical practice and eroded the foundations of general practice (Competition Commission South Africa, 2019; Coovadia et al., 2009). General practitioners have ceded traditional responsibilities to specialists, patients struggle to navigate an increasingly complex system, and the coordination that characterises effective healthcare has become the exception rather than the norm. This fragmentation carries significant costs. Patients with chronic conditions receive episodic rather than continuous care. Preventable hospital admissions occur with regularity. Healthcare expenditure concentrates in high-cost settings while primary care remains underfunded (Competition Commission South Africa, 2019). This is compounded by the difficulty patients experience in navigating health literacy and resources to coordinate their own care (Greenhalgh, 2009; Lall et al., 2018).

The gap

International evidence consistently demonstrates that integrated, team-based primary care is associated with better outcomes and, in some settings, lower costs than fragmented fee-for-service models (Kringos, 2025; Kringos et al., 2013; Starfield et al., 2005). The Patient-Centred Medical Home, integrated care systems in Europe, and comprehensive primary care organisations internationally point towards common principles: care coordination, population health management, multidisciplinary teamwork, and alignment between payment and outcomes (Bodenheimer et al., 2014; Curry & Ham, 2010). Yet translating these principles into established fee-for-service environments has proven persistently difficult. Clinicians trained in individualistic practice resist team-based approaches. Payment systems reward volume rather than value. Information systems do not support coordination. The organisational infrastructure needed for transformation (coaching, quality improvement and data analytics) is absent from private primary care (Bodenheimer et al., 2014; Curry & Ham, 2010).

Paper purpose

This paper describes a decade of developing and implementing an integrated primary care hub model in South Africa. We present a practice account: lessons learned through implementation. We focus on three questions: what worked, what did not, and what our experience offers for the health system reforms underway (including National Health Insurance [NHI]), as South Africa strives towards universal health coverage (UHC).

CONTEXT

South African healthcare landscape

South Africa operates a two-tiered healthcare system characterised by stark inequalities. Medical schemes cover approximately 9.13 million beneficiaries, around 15% of the population, while the public sector serves the remaining majority (Council for Medical Schemes, 2025a). Despite covering a minority, private-sector delivery has historically accounted for approximately half of total healthcare expenditure (Competition Commission South Africa, 2019; Coovadia et al., 2009). Private healthcare is funded primarily through medical schemes: not-for-profit health insurance entities regulated by the Council for Medical Schemes (Competition Commission South Africa, 2019).

The private sector has historically been hospital-centric. Benefit designs have provided comprehensive cover for inpatient care while limiting outpatient benefits, including through mechanisms such as medical savings accounts that shift costs to members (Competition Commission South Africa, 2019). This structure created predictable responses: clinicians delivered care in hospital settings where reimbursement was assured, while patients navigated an inpatient-driven supply chain even for conditions better managed in primary care (Competition Commission South Africa, 2019). General practitioners (GPs) found themselves increasingly marginalised as specialists absorbed functions they traditionally performed. The conditions for comprehensive general practice have eroded, with implications for the skills, confidence, and systems knowledge GPs need to coordinate care effectively. These dynamics are not accidental. Private primary care in South Africa is fragmented by design, through five mutually reinforcing features: concentration of power among a small number of hospital groups and funders; perverse incentives rewarding volume over outcomes; passive purchasing by schemes that reimburse rather than contract for value; asymmetric information favouring suppliers over patients; and limited transparency (Competition Commission South Africa, 2019). Each feature reinforces the others, producing a stable equilibrium of high expenditure, hospital dependence, and weak primary care.

Policy environment

South Africa has committed to universal health coverage through National Health Insurance, which was passed by Parliament in 2023 and signed into law in 2024 (Republic of South Africa, 2024). The NHI vision emphasises primary healthcare as the foundation of the system, with GPs playing a central coordinating role. However, implementing this vision requires transforming both supply and demand sides of healthcare delivery (National Department of Health, 2024). The policy environment creates both pressure and opportunity for primary care innovation. Medical schemes face incentives to demonstrate value and sustainability (Kasapato, 2025). The healthcare workforce, while entrenched in fee-for-service practice, increasingly recognises that current arrangements are unsustainable.

CONCEPTUAL FRAMEWORK

Informing principles

The Chronic Care Model articulated by Wagner et al. (2001) identifies key elements for effective management of chronic conditions: self-management support, delivery system design, decision support, clinical information systems, and community resources. The Patient-Centred Medical Home concept emphasises comprehensive first-contact care, coordination across settings, and quality improvement (Agency for Healthcare Research and Quality, 2022). Evidence from medical home implementations shows that practice transformation requires sustained support, not merely training or incentives (Kringos, 2025; Nutting et al., 2009). Kaiser Permanente's integrated model demonstrates how payer–provider integration can enable population health management at scale (Feachem et al., 2002; Kringos, 2025). Three terms recur throughout this paper and merit brief operational definition. Integration refers to coordinated delivery of services across providers, settings and time, organised around the patient rather than the provider. Care coordination is the deliberate organisation of patient-care activities between two or more participants involved in a patient's care, with the goal of facilitating the appropriate delivery of healthcare services. Practice transformation describes the sustained change in how a general practice organises its team, workflows and patient relationships in order to deliver integrated, team-based care.

Adapting for context

The South African context differs from settings where most integrated care research originates. The fee-for-service environment is deeply entrenched, with limited experience of alternative payment models. General practice operates as isolated small businesses rather than organised provider groups. The infrastructure for

care coordination (shared information systems, care management protocols, quality improvement capacity) is absent. Any model seeking to implement integrated care must address these structural realities. The absence of organisational infrastructure meant that capabilities taken for granted elsewhere (quality improvement, care management, data analytics) had to be built from scratch. This required an organisation positioned between funder and provider that could supply enabling functions without requiring individual practices to develop them independently. The fee-for-service mindset meant practice change would require more than altered incentives. Clinicians needed support to develop new skills, confidence in new ways of working, and evidence that change was possible. This pointed towards coaching and facilitated learning rather than directive management. Crucially, we recognised that putting doctors in charge required acknowledging that most had no experience managing delivery systems. Three decades of a fragmented private primary care system had taught clinicians to stay in their lane, not thinking systemically about population health. Transforming clinical practice would require sustained support, which we term practice transformation coaching.

Core design principles

From this analysis emerged several non-negotiable design principles. The general practitioner would remain the clinical leader and custodian of their patient population: we sought to centre and support GPs, not replace them. Integration would be built through multidisciplinary teamwork, with team members sharing knowledge of patients and coordinating their efforts, an approach consistent with the value-based geography model of care (Porter et al., 2020). Data would drive improvement, with risk stratification enabling proper matching of patient needs to service intensity. And payment would align with value, creating financial conditions conducive to quality and efficiency rather than volume.

THE HUB MODEL

Overview

The integrated care hub model organises participating general practices into coach-led local multidisciplinary teams supported by centralised enabling functions. Each hub serves as the operational unit through which programme principles translate into daily practice, enabling coordinated, proactive care rather than fragmented episodic activity. The model operates at three levels: the individual practice where clinical care is delivered; the hub where practices collaborate and learn together; and the central organisation which provides enabling infrastructure.

Hub structure

Each hub includes a cluster of general practices within a coherent geographic area, led by a Practice Transformation Coach who manages both human and technical aspects of change (see Figure 1). Care Coordinators are clinical staff (nurses and clinical associates) deployed to practices based on a dynamic capacity model matching staffing to population need, considering enrolled population size, disease burden, and expected workload. The multidisciplinary team extends beyond the practice to include allied health professionals: physiotherapists, occupational therapists, dietitians, social workers, and registered counsellors, who together provide services traditionally largely unavailable in outpatient primary care due to benefit limitations. Under the value-based contract, these services become accessible to patients who need them, enabling conservative management that can prevent or delay costly interventions. Integration within the hub is horizontal: primary care colleagues from complementary disciplines work together.

The monthly hub multidisciplinary team meeting, chaired by the Coach and enabled with technology for shared clinical information and continuity of care, serves as the forum for case discussion, team learning, and continuous improvement.

The coach role

The Practice Transformation Coach plays the central role in enabling practice change, establishing the shared mindset required for value-based primary care, consistently reinforcing workflows that enable proactive team-based care, and supporting clinicians as they develop new ways of working. Coaches require both emotional intelligence (capacity to shift clinical mindsets and build trust) and technical intelligence, including familiarity with programme tools, care pathways, and data for improvement. The coaching model recognises that transformation cannot be achieved through training alone; changing behaviour embedded over decades requires sustained support and trusted relationships. Coaches do not manage clinical decisions but create conditions in which effective team-based care can flourish.

Central enabling functions

While hubs deliver care locally, they depend on centralised functions that no individual practice could efficiently provide: legal experts managing funder contracts; human resources managing care coordinator recruitment; a clinical quality unit designing training and care pathways; analytics producing performance reports; and an intelligent IT system providing digital infrastructure supporting care continuity (see Figure 2). This platform maintains patient records accessible to all team members, generates task workflows, enables risk stratification based on claims, clinical, and psychosocial data, and produces operational and outcome reports. The technology enables but does not drive transformation; it is a necessary but not sufficient condition for integrated care.

Population health management

The model takes a population health approach, segmenting enrolled patients by risk level and matching service intensity to need. Healthy members receive prevention and screening services; those at risk receive monitoring; patients with chronic conditions receive care plan support; complex patients receive intensive care coordination with frequent multidisciplinary team involvement. Whole-patient assessments capture not only clinical information but psychosocial factors (housing, income, social support, functional status) that influence health outcomes and care needs; the underlying data architecture is described in a companion article (Knight et al., 2026).

Evolution over time

The model described here did not appear fully formed. The first pilot site launched in 2016, organised around a small group of committed general practitioners willing to work differently. Lessons from that pilot informed a second site in 2019, which proved the approach could be replicated. These pilots eventually led to a national population medicine contract with a large restricted medical scheme in 2023. Through this journey, every element of the model has evolved: coaching approaches have been refined, technology platforms rebuilt, payment models adjusted and care pathways revised. The model continues to evolve.

OPERATIONAL EXPERIENCE

Operationally, the programme organises around three phases: Build, Enable, and Operate (see Figure 3). The build phase identifies and contracts suitable GP practices against value-based principles and establishes the financial and system readiness required. The enable phase staffs the hub with a core local team (Practice Transformation Coach, care coordinators, and allied health professionals) scaled to patient load and develops partnerships with local clinical and social services. The operate phase runs the hub for clinical excellence: population risk assessment triggers care plans with task execution through multidisciplinary coordination and learning-oriented MDT meetings. Central services (finance, HR, IT and reporting) provide the scaffolding that enables local teams to focus on care delivery.

What worked

The value-based contract covered varying percentages of participating GP practices' patient populations depending on their exposure to the medical scheme, the third-party funder. Some practices had well over half their patients in the programme, while others had approximately 10% participating. The ability of practices to adopt all programme requirements varied accordingly. Several elements proved essential. First, investment in coaching was essential, as clinicians needed ongoing support to transform practice workflows embedded over their careers. Coaches combining clinical credibility with change management skills became critical enablers. The monthly hub meetings became forums for building shared identity and trust. Second, a focus on the general practitioner as the clinical leader was vital; early attempts to build systems around care coordinators failed. GPs who felt marginalised disengaged; in contrast, those who felt supported flourished. Third, the discipline of data-driven improvement with regular performance reports shared transparently created accountability and opportunity. Fourth, integration of allied health professionals, whose services were previously unavailable, enabled conservative management.

What did not work

First, early pilots treated hubs as businesses: registered as proprietary limited companies with practices as shareholders. In practice, doctors' ability to manage staff and play required managerial roles was limited; the organisation had to step in, resulting in a rollback of this model. This was replaced with management centred within our organisation and a strong structured GP support system. Second, technology implementations prioritising features over usability frustrated clinicians. Third, the assumption that patients would eagerly embrace proactive care coordination proved naïve; many accustomed to episodic healthcare initially resisted ongoing engagement. Lastly, we also underestimated the difficulty of unlearning entrenched clinical practice; clinicians trained over decades in fee-for-service approaches do not transform overnight.

Critical success factors

From a decade of experience, several factors emerge as critical. First, clinical champions: GPs willing to model different practice proved essential in every successful hub. Second, sustained organisational commitment: building integrated care requires years of investment before returns materialise. Third, appropriate payment: fee-for-service and integrated care are incompatible; value-based payment creates necessary conditions. Fourth, structural investment in the central enabling functions (analytics, quality improvement, care coordination infrastructure) without which local hubs cannot operate at scale.

RESULTS OBSERVED

Measurement approach

Assessing the impact of complex interventions in real-world settings presents well-known challenges. Randomised controlled trials are not feasible; selection effects and confounding factors limit causal inference. We present available evidence with appropriate caution, as indicated below. Utilisation outcomes (medical admission rate, bed days, surgical admission rate) are reported against a blended benchmark: a risk-adjusted expected rate for the comparable population, derived from a combined national and local peer-provider reference used in contract-level outcome analysis. Process measures (vaccination, HbA1c testing) are reported against the cohort's pre-programme baseline rate. Although the programme has been documented in an externally commissioned international policy case study (Smith & Mosam, 2023), no formal independent comparative evaluation has been conducted, and causal attribution to the programme intervention remains an active area of methodological development, with emerging frameworks advancing what is possible in real-world evidence (McClellan et al., 2026; Skivington et al., 2021). Our measurement

approach combines operational metrics with outcome indicators defined in our contract with the funding scheme.

Programme scale

The programme grew substantially from a first pilot site in 2016 serving several hundred members to a national programme. Enrolled populations expanded from 2,300 members at the start of the 2019 pilot (5,620 by mid-2022; Smith & Mosam, 2023) to over 135,000 members across five provinces by 2025. Participating GP practices grew from a handful in pilot projects to over 230 organised into 15 hubs, supported by more than 55 care coordinators and 11 Practice Transformation Coaches. Enrolled patients are stratified by risk: approximately 65% healthy, 20% at-risk, 13% with chronic conditions, and 2% complex patients with multimorbidity (Internal programme data, 2023–2025).

Clinical outcomes

Subject to the limitations noted above, internal programme evaluation found patterns consistent with programme objectives. Analysis against risk-adjusted benchmarks for the enrolled population of the 2019 pilot site over the first three quarters of 2021 (Table 1) found that the medical admission rate was 29.6% lower than expected. The overall number of patient bed days was 7% below predicted levels. The surgical admission rate, however, was 20.5% higher than expected (Table 1). This result may reflect conditions specific to the pilot area at the time, including an oversupply of surgical hospital capacity and surgical specialist density, compounded by a backlog of elective cases delayed by the COVID-19 pandemic (Internal programme evaluation, 2021). The pattern is most plausibly explained by incomplete control of these comparator factors rather than by a hub-induced increase in surgical referrals; this attribution remains a candidate hypothesis for future evaluation.

Process measures showed improvements in preventive care. Influenza vaccination uptake among patients with chronic conditions or aged over 65 increased by 39%. HbA1c testing rates among diabetic patients showed a 5% improvement. Aggregated performance quality scores showed positive trends over successive measurement periods (Internal programme data, 2023–2025).

Programme sustainability

The initial funding scheme has continued and expanded the programme over successive contract periods, which may be interpreted as indicative of perceived value. A second medical scheme has announced intent to join. These decisions suggest the model delivers value sufficient to justify continued investment.

DISCUSSION

Implications for practice

Our experience has several implications. First, expect transformation to take years: clinical practice that produces fragmentation developed over decades and will not change quickly. Managers, organisations and funders must commit to sustained investment. Second, invest heavily in human elements: technology enables but does not drive transformation; coaching, relationships, and trust prove more important than systems (Gittell et al., 2013; Lee, 2025). Building and nurturing coaching capability deserves at least as much attention as building technical capability. Third, work with clinical champions and let success spread: mandating participation produces compliance at best, resistance more often. Movement building proves more effective than programme management.

Policy relevance

This experience offers evidence that integrated primary care can work within private sector structures in South Africa when funders act as strategic purchasers willing to commission restructuring and buy value. This matters for NHI implementation, which envisions purchasing integrated care from both public and private providers. More broadly, the experience also speaks to middle-income countries seeking to strengthen primary care; the challenges of fragmented systems, isolated practices, and limited infrastructure are common across many settings. The economics of this restructuring are compelling. In 2024, of every R100 in medical scheme benefits paid, R35.95 went to hospital services and R28.02 to specialists (much of whose work occurs in-hospital), while only R5.13 reached general practitioners and R8.47 reached allied health professionals (Council for Medical Schemes, 2025b). Much of the hospital-associated activity could be avoidable with stronger primary care (Competition Commission South Africa, 2019). Even a modest rebalancing makes a material difference. A five-percentage-point shift from hospital and specialist spending towards general practice would almost double current GP funding (from R5.13 to R10.13 per R100). Such a shift is achievable by preventing unnecessary admissions and bed days through earlier, better-coordinated outpatient care, and this could translate into system-wide savings without reductions in access or quality. Under population-based payment arrangements, the resulting savings can flow back to providers and members rather than to intermediaries, funding the very capacity that makes further savings possible while curbing waste (James & Poulsen, 2016). These are not efficiencies extracted from existing budgets but the dividend of structural change. For funders, the question shifts from containing hospital costs to investing in the primary care that avoids them.

Limitations and transferability

Several limitations constrain generalisation of our experience. We have operated primarily with a single large medical scheme; serving multiple schemes simultaneously presents additional complexity. No formal external evaluation has been conducted. We operate in South African conditions that differ from other contexts in ways that may matter for transferability. However, we believe the principles underlying our approach (clinical leadership, coaching for transformation, aligned payment, data-driven improvement and multidisciplinary teamwork) are likely to apply across contexts, even if specific implementation details require local adaptation. Formal external quasi-experimental evaluation of this programme is a clear next step. Comparable evaluations of integrated and comprehensive primary care models in other health systems, including difference-in-differences analyses of the UK National Health Service Vanguard Models programme (Morciano et al., 2020) and the United States Comprehensive Primary Care Plus initiative (Singh et al., 2024), offer methodological precedents that could be adapted to the South African private-sector context. Applied to this programme, such an approach may allow the descriptive outcome indicators reported above to be quantified against appropriate comparison groups, with adjustment for confounding and measures of uncertainty.

CONCLUSION

A decade of implementing integrated primary care hubs in South Africa's private healthcare sector reinforces three fundamental lessons. Transformation requires sustained coaching, not just training. Technology enables but does not drive change. Value-based payment creates necessary conditions but cannot substitute for building teams. Two supporting lessons follow: the general practitioner must remain the clinical leader and custodian of patients, and clinical champions, supported and celebrated, become catalysts for broader transformation. When we began this work, we were told South African private healthcare was too entrenched to change, that fee-for-service practice was immutable, that general practitioners could not lead teams. Our experience suggests otherwise. With appropriate support, clinicians can transform their practice. With aligned incentives, teams can form and flourish. With sustained commitment, fragmented systems can

integrate. The work is hard and progress slow, but the destination, healthcare that is coordinated, proactive and centred on patients rather than transactions, is worth the journey.

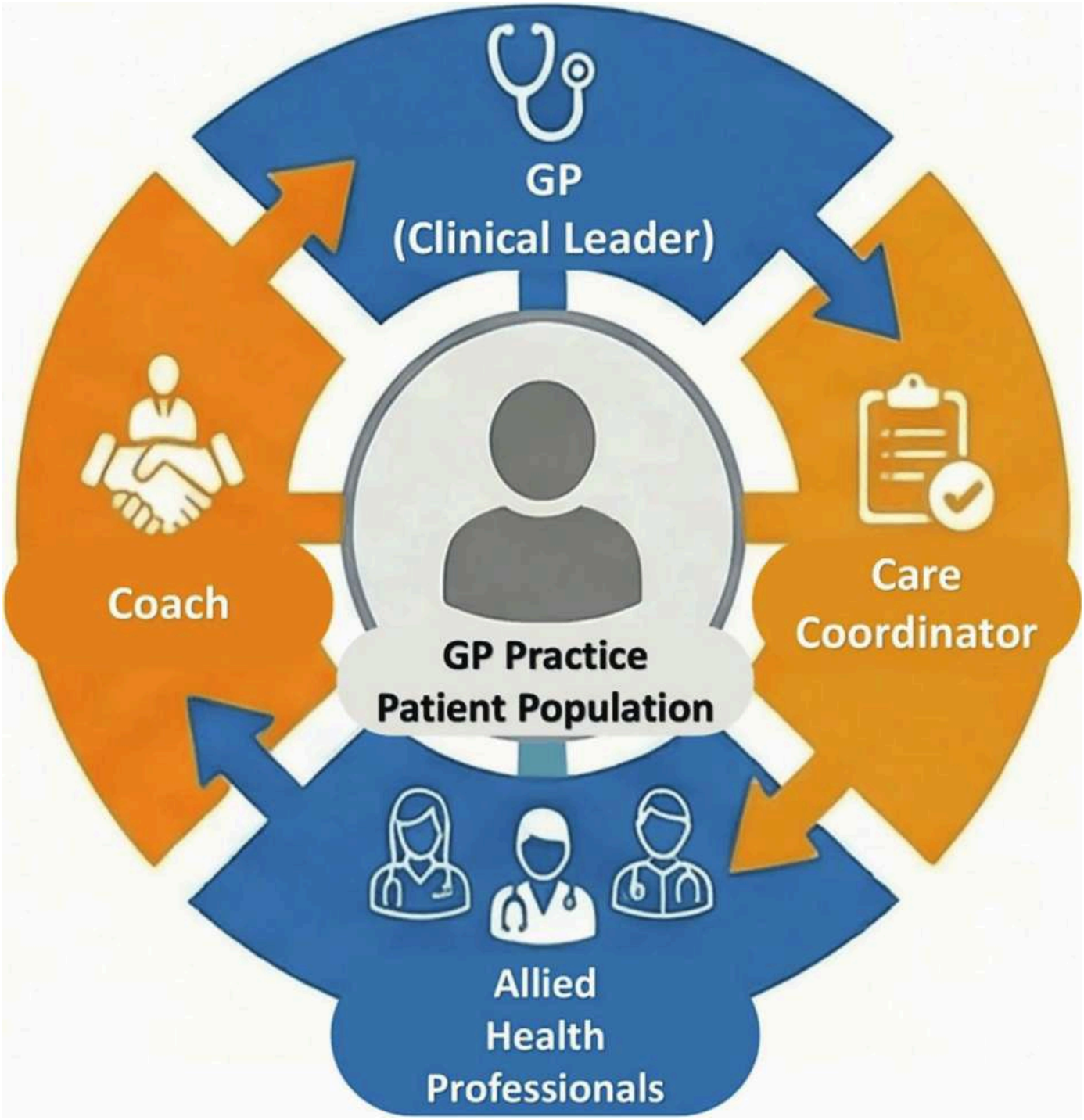

Figure 1: Composition of the GP-led local multidisciplinary team

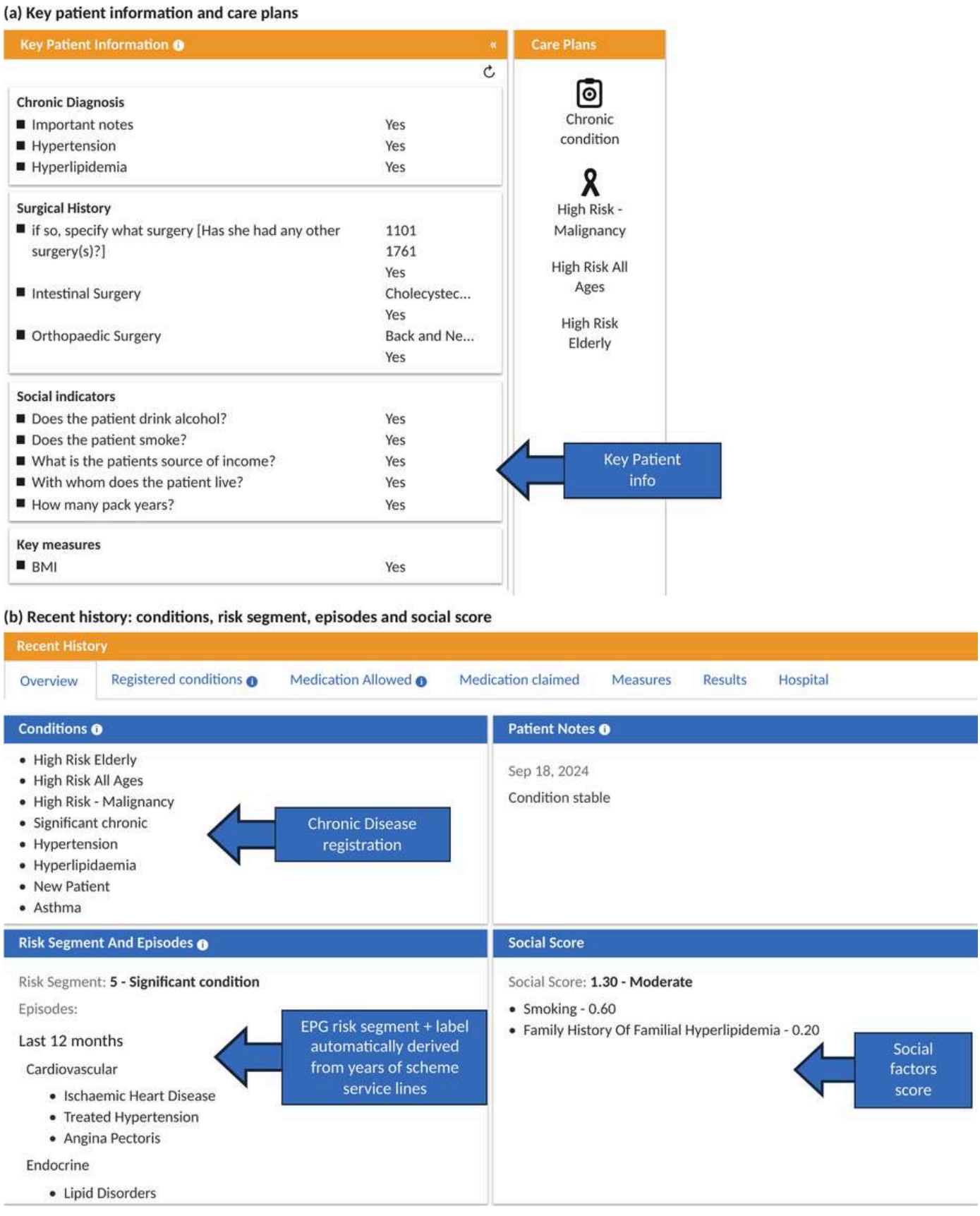

Figure 2: Intelligent IT system: (a) key patient information and care plans; (b) recent history, showing automated risk stratification and the whole-patient view that supports multidisciplinary team workflow

Note: EPG = Episode Population Grouper. The record shown is de-identified and contains no direct patient identifiers.

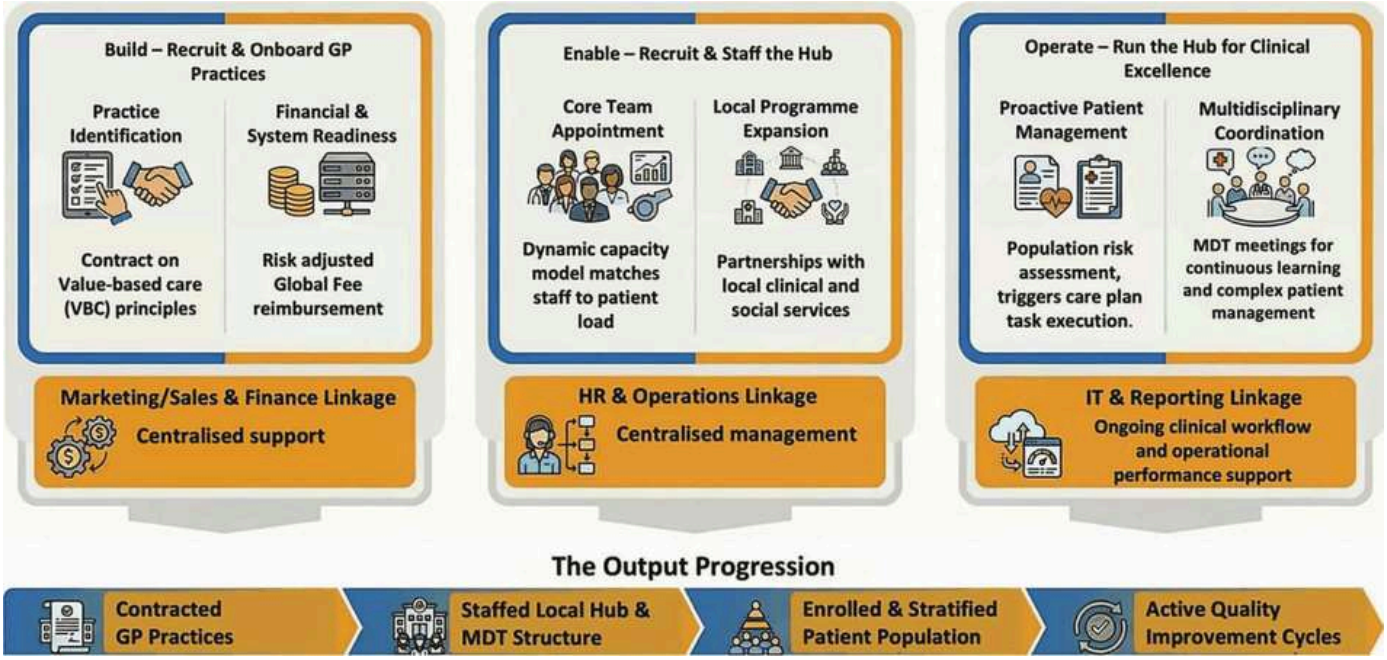

Figure 3: Build–Enable–Operate phases of hub establishment and operation

Table 1: Summary of key outcome indicators

Indicator	Result	Benchmark
Medical admission rate	29.6% lower	Blended benchmark†
Patient bed days	7% lower	Blended benchmark†
Flu vaccination (chronic/65+)	39% increase	Baseline rate‡
HbA1c testing (diabetics)	5% increase	Baseline rate‡
Surgical admission rate	20.5% higher*	Blended benchmark†

*May reflect surgical hospital oversupply and the COVID-19 backlog. †Blended benchmark: a risk-adjusted expected rate for the comparable population, derived from a combined national and local peer-provider reference used in contract-level outcome analysis. ‡Baseline rate: the cohort's pre-programme rate for the same indicator. Sources: utilisation indicators (medical admissions, bed days, surgical admissions), internal programme evaluation, first three quarters of 2021 (also reported in Smith & Mosam, 2023); process indicators (influenza vaccination, HbA1c testing), internal programme data, 2023–2025.

REFERENCE LIST

Agency for Healthcare Research and Quality. (2022). *Defining the PCMH*. https://www.ahrq.gov/ncepcr/research/care-coordination/pcmh/define.html

Bodenheimer, T., Ghorob, A., Willard-Grace, R., & Grumbach, K. (2014). The 10 building blocks of high-performing primary care. *Annals of Family Medicine*, *12*(2), 166–171. https://doi.org/10.1370/afm.1616

Competition Commission South Africa. (2019). *Health Market Inquiry: Final findings and recommendations report*.

Coovadia, H., Jewkes, R., Barron, P., Sanders, D., & McIntyre, D. (2009). The health and health system of South Africa: Historical roots of current public health challenges. *The Lancet*, *374*(9692), 817–834. https://doi.org/10.1016/S0140-6736(09)60951-X

Council for Medical Schemes. (2025a). *Annual report 2024/25*. https://www.medicalschemes.co.za/cms-annual-report-2024-25/

Council for Medical Schemes. (2025b). *The medical schemes industry in 2024*. https://www.medicalschemes.co.za/wp-content/CMSIndustryReport2024_4Dec.pdf

Curry, N., & Ham, C. (2010). *Clinical and service integration: The route to improved outcomes*. The King's Fund. https://www.kingsfund.org.uk/insight-and-analysis/reports/clinical-and-service-integration

Feachem, R. G., Sekhri, N. K., & White, K. L. (2002). Getting more for their dollar: A comparison of the NHS with California's Kaiser Permanente. *BMJ*, *324*(7330), 135–141. https://doi.org/10.1136/bmj.324.7330.135

Gittell, J. H., Godfrey, M., & Thistlethwaite, J. (2013). Interprofessional collaborative practice and relational coordination: Improving healthcare through relationships. *Journal of Interprofessional Care*, *27*(3), 210–213. https://doi.org/10.3109/13561820.2012.730564

Greenhalgh, T. (2009). Patient and public involvement in chronic illness: Beyond the expert patient. *BMJ*, *338*, Article b49. https://doi.org/10.1136/bmj.b49

Internal programme data. (2023–2025). *Operational metrics, outcome indicators, and performance reports* [Unpublished internal data]. Professional Provider Organisation Services (Pty) Ltd.

Internal programme evaluation. (2021). *Outcome analysis against risk-adjusted benchmarks* [Unpublished internal report]. Professional Provider Organisation Services (Pty) Ltd.

James, B. C., & Poulsen, G. P. (2016). The case for capitation: It's the only way to cut waste while improving quality. *Harvard Business Review*, *94*(7–8), 102–111.

Kasapato, L. (2025). Exploring the future role of medical schemes in South Africa's evolving UHC/NHI landscape: From passive payers to strategic partners. *Southern African Health Journal*, 6th ed., 20–29. https://www.bhfglobal.com/Publication?report=6617107c-435b-f111-bec6-7ced8d5c37e0

Knight, S., Ruff, B., Jabaar, R., & Chikezie, P. (2026). Data for value: Integrating clinical and social risk data for population health management. *Council for Medical Schemes Research Journal*, *1*(1), 122–136.

Kringos, D. S. (2025). Primary care. In *The Wiley Blackwell encyclopedia of health, illness, behavior, and society* (2nd ed.). Wiley. https://doi.org/10.1002/9781118410868.wbehibs384.pub2

Kringos, D., Boerma, W., Bourgueil, Y., Cartier, T., Dedeu, T., Hasvold, T., Hutchinson, A., Lember, M., Oleszczyk, M., Rotar Pavlic, D., Svab, I., Tedeschi, P., Wilm, S., Wilson, A., Windak, A., Van der Zee, J., & Groenewegen, P. (2013). The strength of primary care in Europe: An international comparative study. *British Journal of General Practice*, *63*(616), e742–e750. https://doi.org/10.3399/bjgp13X674422

Lall, D., Engel, N., Devadasan, N., Horstman, K., & Criel, B. (2018). Models of care for chronic conditions in low/middle-income countries: A 'best fit' framework synthesis. *BMJ Global Health*, *3*(6), Article e001077. https://doi.org/10.1136/bmjgh-2018-001077

Lee, T. H. (2025). *Social capital in healthcare: How trust and teamwork drive organizational excellence*. Wiley.

McClellan, M., Pegus, C., Ford, A., Armstrong, C., & Rego, E. (2026). *Health Value Evidence Quality: A framework to advance private investment that improves health care and health*. Duke-Margolis Institute for Health Policy, Duke University. https://healthpolicy.duke.edu/publications/capital-impact-council-health-value-evidence-quality-framework

Morciano, M., Checkland, K., Billings, J., Coleman, A., Stokes, J., Tallack, C., & Sutton, M. (2020). New integrated care models in England associated with small reduction in hospital admissions in longer-term: A difference-in-differences analysis. *Health Policy*, *124*(8), 826–833. https://doi.org/10.1016/j.healthpol.2020.06.004

National Department of Health. (2024). *National Health Insurance*. https://www.health.gov.za/nhi/

Nutting, P. A., Miller, W. L., Crabtree, B. F., Jaén, C. R., Stewart, E. E., & Stange, K. C. (2009). Initial lessons from the first national demonstration project on practice transformation to a patient-centered medical home. *Annals of Family Medicine*, *7*(3), 254–260. https://doi.org/10.1370/afm.1002

Porter, M. E., Lee, T. H., & Murray, A. C. A. (2020). The value-based geography model of care. *NEJM Catalyst Innovations in Care Delivery*, *1*(2). https://doi.org/10.1056/CAT.19.1130

Republic of South Africa. (2024). *National Health Insurance Act 20 of 2023* (Government Gazette No. 50664, 16 May 2024). Government Printer. https://www.gov.za/documents/acts/national-health-insurance-act-20-2023-english-afrikaans-16-may-2024

Singh, P., Fu, N., Dale, S., Orzol, S., Laird, J., Markovitz, A., Shin, E., O'Malley, A. S., McCall, N., & Day, T. J. (2024). The Comprehensive Primary Care Plus model and health care spending, service use, and quality. *JAMA*, *331*(2), 132–146. https://doi.org/10.1001/jama.2023.24712

Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M., & Moore, L. (2021). A new framework for developing and evaluating complex interventions: Update of Medical Research Council guidance. *BMJ*, *374*, Article n2061. https://doi.org/10.1136/bmj.n2061

Smith, A., & Mosam, A. (2023). *Value Care Team model in South Africa: Case study*. World Health Organization & Organisation for Economic Co-operation and Development. https://iris.who.int/handle/10665/373212

Starfield, B., Shi, L., & Macinko, J. (2005). Contribution of primary care to health systems and health. *The Milbank Quarterly*, *83*(3), 457–502. https://doi.org/10.1111/j.1468-0009.2005.00409.x

Wagner, E. H., Austin, B. T., Davis, C., Hindmarsh, M., Schaefer, J., & Bonomi, A. (2001). Improving chronic illness care: Translating evidence into action. *Health Affairs*, *20*(6), 64–78. https://doi.org/10.1377/hlthaff.20.6.64

DECLARATIONS

Acknowledgements

The authors wish to acknowledge the general practitioners, care coordinators, allied health professionals, and management team whose work over more than a decade has made the practice described in this paper possible. We are grateful to the medical scheme partner whose contracting framework enabled the programme. No external research funding was received for the preparation of this manuscript; the underlying programme of work has been funded through value-based contracts between the authors' organisation and a major South African medical scheme.

Author contributions

Hellen Nkwagatse: Conceptualisation, Methodology, Writing – Original Draft.

Belinda Richards: Conceptualisation, Methodology, Supervision, Writing – Review & Editing.

All authors have read and approved the final version of this manuscript.

Data availability statement

The aggregated, de-identified operational data presented in this paper are derived from internal programme evaluation and are the property of the authors' organisation and its contracting medical scheme partner. Requests for further data access may be directed to the corresponding author, subject to institutional and contractual data-sharing constraints.

Declaration of competing interest

The authors are employed by Professional Provider Organisation Services (Pty) Ltd, the organisation that designed and delivers the integrated primary care programme described in this paper. As is inherent in a value-based contract, part of the organisation's contract income is linked to performance on contracted utilisation and quality indicators; the authors' individual remuneration is not determined by the specific indicators reported in this article. No other financial conflicts of interest or personal relationships that could have influenced the work are declared.

Declaration of ethics

The authors declare that the writing of this manuscript adheres to ethical standards of publication, in accordance with scientific principles, and is free from plagiarism. The programme described in this paper is delivered by clinicians and an organisation that hold direct ethical responsibility for the care of enrolled patients, operating within the clinical governance frameworks of the Health Professions Council of South Africa (HPCSA) and the professional codes that bind practitioners to the care of their patients in South Africa. This paper is a practice account of that operational programme rather than a research study involving identifiable patient data or human research participants. No research ethics committee approval or waiver was sought: the data reported are aggregated, de-identified operational and contract-level indicators generated in the course of routine service delivery and evaluation, no individual-level data are reported, and the record shown in Figure 2 is de-identified.

Declaration of assistive technologies in the writing process

The authors declare that generative artificial intelligence (GenAI) and AI-assisted tools were used in a limited and responsible manner during the preparation of this manuscript. Specifically, large language models were used for drafting assistance, structural editing, and reference verification. All AI-generated content was critically reviewed, revised, and validated by the authors against original sources to ensure accuracy, originality, and compliance with ethical and scholarly standards. All technical content, analytical interpretations and conclusions are the work of the authors. The authors take full responsibility for the final content of this manuscript.



STREAM I
VALUE-BASED CARE MODELS

TO WHAT EXTENT CAN NURSE-LED, TEAM-BASED PHC DELIVERY MODELS IMPROVE COST EFFICIENCY, ACCESS, AND PROVIDER INCENTIVES RELATIVE TO TRADITIONAL SOLO GP MODELS IN SOUTH AFRICA? A SHARED VALUE ASSESSMENT OF ALTERNATIVE MODELS FOR PRIMARY CARE DELIVERY

Roseanne Harris, Nondumiso Gugu Beatrice Khumalo and Sarika Besesar

How to cite this article: Harris, R., Khumalo, N. G. B., & Besesar, S. (2026). To what extent can nurse-led, team-based PHC delivery models improve cost efficiency, access, and provider incentives relative to traditional solo GP models in South Africa? A shared value assessment of alternative models for primary care delivery. *Council for Medical Schemes Research Journal*, *1*(1), 39–58.

ABSTRACT

Background: Primary care systems face increasing pressure to expand access, improve quality and contain costs within constrained resources. This study evaluates whether a nurse-led, team-based primary health care delivery model can improve efficiency, patient access and provider incentives relative to traditional solo general practitioner practice models.

Methods: A comparative modelling approach was applied using a representative South African solo GP practice as the baseline, assuming approximately 1,500 patients and standard consultation patterns. The alternative model introduced nurse-led triage within a multidisciplinary team, with defined task-sharing, referral pathways, supervision requirements and differentiated remuneration levels. Scenario analyses assessed impacts on service capacity, cost efficiency, GP workload and financial performance.

Results: The model indicates substantial gains across all dimensions. Patient coverage increases by approximately 3.8 times, driven by nurse-led management of lower-complexity cases. Assuming a fee per visit about 36% lower (R350 rather than R550), net practice revenue to the GP still increases by about 41%, driven by higher throughput and productivity. The model enables better alignment of provider time with patient complexity, allowing more focused management of complex cases.

Conclusions: These findings suggest that team-based PHC could be a scalable and financially sustainable approach to strengthening primary care. In South Africa, the model could expand affordable access, reduce out-of-pocket expenditure and support low-cost benefit design aligned with universal health coverage and National Health Insurance priorities, while improving workforce utilisation and enabling job creation in the health sector. Implications also include stronger system resilience and improved quality through structured clinical pathways and enhanced care coordination across multidisciplinary teams.

KEYWORDS

Primary health care; team-based care; nurse-led triage; health system efficiency; universal health coverage; cost optimisation

INTRODUCTION

Universal primary healthcare is defined as essential and affordable care that is accessible to everyone, and includes health promotion, disease prevention, health maintenance, education and rehabilitation. Since the 1978 Declaration of Alma-Ata, which was reinforced by further resolutions of the World Health Assembly, the World Health Organization (WHO) has promoted primary healthcare as a core component of health systems (WHO & United Nations Children's Fund [UNICEF], 2018). Evidence has shown that a focus on primary and preventive care has the effect of reducing the need for more expensive tertiary care, thus improving efficiencies, effectiveness and outcomes (including patient satisfaction) of the healthcare system. Primary health care (PHC) has proven to be a highly effective and efficient way to address the main causes of and risk factors for poor health, as well as for handling the emerging challenges that may threaten health in the future (WHO & UNICEF, 2018). This also has additional socioeconomic benefits that would accrue from improving the health status and productivity of the entire population and greater socioeconomic equity. For example, the WHO notes that primary healthcare interventions across low- and middle-income countries (LMICs) could save up to 60 million lives and increase average life expectancy by 3.7 years by 2030 (WHO & UNICEF, 2018).

Similarly, the concept of universal health coverage (UHC), as noted in the United Nations' 2015 Sustainable Development Goals, is an aspiration to provide all people with access to essential high-quality health services while ensuring financial risk protection by providing care regardless of a person's ability to pay for it (WHO, 2019b).

The UHC model is also based on PHC as the first point of access into the health system. From these two definitions it is clear that there is overlap between the aims of PHC and UHC. Many have argued that to achieve UHC, reforms should therefore focus on strengthening primary health care (van Weel & Kidd, 2018). In particular, as noted in the WHO's Sustainable Development Report, the UHC service coverage index combines multiple tracer indicators of service coverage into a single summary measure and is strongly influenced by access to essential PHC services (WHO, 2019b). The metrics include access to family planning services, antenatal care, childhood immunisations and HIV treatment services, among other indicators used to assess progress towards universal health coverage (WHO, 2019b).

LITERATURE REVIEW

The Lancet Global Health Commission on financing primary health care (Hanson et al., 2022) noted that provider payment mechanisms in many low- and middle-income countries continue to rely on user fees, input-based budgets, and fee-for-service arrangements, each of which can generate inefficiencies or misaligned incentives. Value-based models, more common in high-income countries, offer predictable funding and promote equity. The Commission recommends blended funding and payment approaches as the foundation for promoting equity, quality and responsiveness for primary health services (Hanson et al., 2022; Mathauer & Wittenbecher, 2013).

According to the Lancet Global Health publication, strengthening PHC is essential for achieving UHC, following earlier evidence that poor-quality care now causes more global deaths than lack of access alone (Hanson et al., 2022; Kruk et al., 2018). The Commission reiterates the importance of the 2018 Astana Declaration's framing of PHC as the most inclusive and efficient system for addressing rising non-communicable diseases and improving overall population health (Hanson et al., 2022; WHO, 2019a). The literature suggests that advancements in PHC financing depend, among other things, on integrating insights from health economics, health systems research and global health, as well as on the role that the private sector can play in filling gaps in public provision (Hanson et al., 2022; Ogunbekun et al., 1999).

The WHO has noted that government funding for PHC in LMICs is significantly below what is required for effective service delivery. In 2018, average public PHC spending was approximately US$3 per capita in low-income countries and US$16 per capita in lower-middle-income countries, far short of WHO's recommended US$65 and US$59, respectively (Hanson et al., 2022; WHO, 2021). In this regard, strengthening PHC delivery and financing is essential for achieving UHC, especially since primary care systems globally face challenges due to rising multimorbidity, workforce shortages, and persistent inequities in access and outcomes. One of the proposed solutions to address these challenges is unlocking innovative nurse-based primary health care clinics. Multidisciplinary team-based care (MTBC) has emerged to address these challenges by redistributing tasks, embedding care coordination, and centring services on patients and communities rather than visits and single professionals (Frikha et al., 2024; Terry et al., 2024).

Nurse-led clinics (NLCs) include structured healthcare services in which advanced practice nurses or nurse practitioners (NPs) deliver clinical assessment, diagnosis, evidence-based treatment, education, and case management, often functioning both independently and interdependently within multidisciplinary pathways (Terry et al., 2024). Within MTBCs, coordinated networks of health professionals, including general practitioners, nurses and nurse practitioners, allied health providers, pharmacists, social workers and behavioural health practitioners, operate with shared goals supported by explicit governance structures, clearly defined roles, and formalised joint operating mechanisms such as case conferences and integrated electronic health records (Frikha et al., 2024; Liang et al., 2025).

Across diverse care settings, care managers, typically nurses or social workers, play a central role in ensuring proactive follow-up, supporting treatment adherence, and facilitating navigation across medical and social services (Frikha et al., 2024). Evidence from the above models demonstrates stronger performance when nurses are enabled to practise to the full extent of their training, including diagnosis, prescribing, test ordering, and leading some chronic disease management activities (within their scope of practice) that are reinforced through standardised protocols and continuous professional development frameworks (Floriancic et al., 2024; Terry et al., 2024). Evidence further suggests that nurse-led clinics are increasingly implemented for chronic disease management, behavioural health, women's and reproductive health, palliative care, and cognitive care for older adults, with relevance in underserved or resource-constrained settings (Sitthimongkol et al., 2025). These models have growing importance for health systems pursuing UHC and equity, where expanding access while preserving quality and safety remains a central reform priority (Frikha et al., 2024; Liang et al., 2025).

STUDY DESIGN

Research question

To what extent can nurse-led, team-based PHC delivery models improve cost efficiency, access, and provider incentives relative to traditional solo GP models in South Africa?

Hypothesis

Null hypothesis (H_0)

There is no significant difference between nurse-led, team-based primary health care (PHC) models and traditional solo GP models in terms of cost efficiency, patient access, and provider remuneration.

Alternative hypothesis (H_1)

Nurse-led, team-based PHC models significantly improve cost efficiency, increase patient access, and maintain or improve provider remuneration relative to traditional solo GP models.

Study aim

To evaluate the impact of a nurse-led, team-based PHC delivery model on service capacity, cost efficiency, and provider remuneration compared to a traditional solo general practitioner (GP) practice model.

Objectives

- To assess the extent to which a team-based PHC model increases patient coverage and service capacity relative to a solo GP practice.
- To evaluate the effect of a nurse-led triage and task-sharing model on the cost of service delivery and cost per patient visit.
- To examine how task redistribution within a multidisciplinary team affects GP workload, time utilisation, and overall productivity.
- To assess whether the team-based PHC model improves financial sustainability and aligns provider incentives with efficient, high-quality care delivery.

Conceptual framework

This study is conceptually grounded in the WHO PHC framework, which emphasises integrated, people-centred health services delivered by multidisciplinary teams as the foundation for achieving UHC. Complementing this, the multidisciplinary team-based care (MTBC) framework provides a delivery-level lens, highlighting the importance of task redistribution, care coordination, and clearly defined team roles in improving system efficiency and access. Within this combined framework, the proposed nurse-led, team-based PHC model represents an operationalisation of PHC principles, enabling the alignment of provider skills with patient complexity, expanding population coverage, and improving cost efficiency while maintaining quality of care (Figure 1).

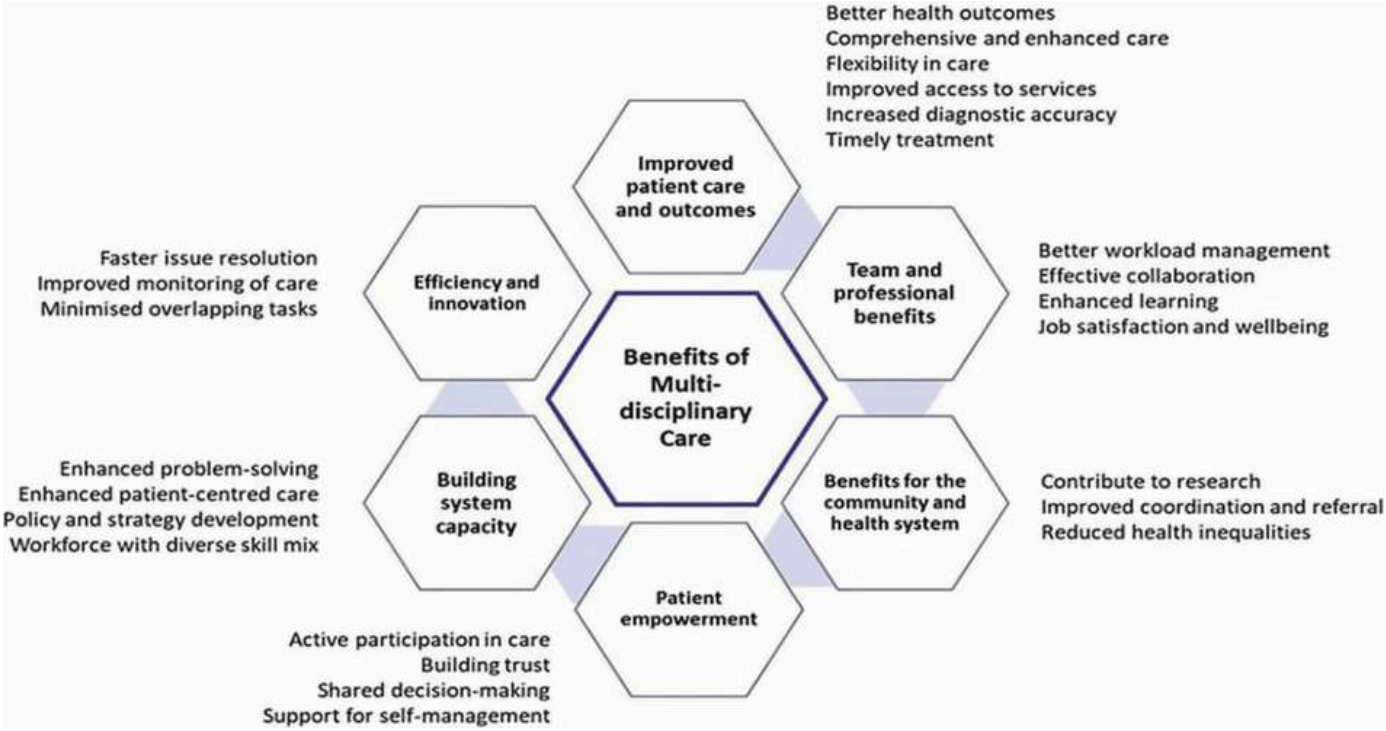

Figure 1: Multidisciplinary team-based care (MTBC) conceptual framework

Source: Adapted from Liang et al. (2025).

METHODOLOGY

Theoretical modelling

In exploring the above framework, the study undertook a hypothetical modelling analysis to evaluate the impact of an alternative primary care model on service capacity, cost efficiency, and remuneration from the perspective of doctors, patients, and other system stakeholders. According to Larrain and Groene (2021), simulation modelling is particularly suited to evaluating complex and integrated health care systems because it can capture the dynamic interactions between system components and assess overall performance in ways that traditional empirical methods cannot. Furthermore, Kim (2024) notes that theoretical modelling approaches enable the analysis of multiple policy-relevant "what-if" scenarios and the estimation of long-term impacts of interventions that are impractical or infeasible to evaluate through experimental or observational study designs.

In addition, Togoveanu (2023) observed that simulation models are central to health system and services evaluation as they integrate diverse data inputs, estimate future costs and outcomes, and explicitly account for uncertainty, thereby supporting robust resource allocation decisions. Together, this literature supports the use of a theoretical modelling approach in this study to rigorously assess the cost, efficiency, access, and overall implications of alternative primary health care delivery models. The analysis relied on a standard solo GP practice as the baseline comparator. Baseline assumptions reflected a representative South African solo practice and included average visit rates, typical consultation durations, and a patient panel of approximately 1,500 unique individuals. Additional parameters incorporated standard staffing configurations, prevailing cost structures, and typical productivity levels observed across solo general practices.

The alternative model introduced a nurse-led triage workflow, designed as a blended care pathway in which nurses manage an initial portion of patient encounters. Assumptions included defined triage and referral rates, GP supervision requirements, and remuneration levels appropriate to each cadre. To assess the implications of this model, we constructed a series of operational and financial scenarios by adjusting discrete levers such as funding flows, reimbursement mechanisms, staffing and capacity configurations, expected changes in patient utilisation, administrative load, technology enablement, and relevant regulatory shifts. We employed a costing model to capture how each scenario would affect practice costs and service delivery. The model integrated cost and revenue components mapped to specific service lines, estimated volume and throughput based on GP time allocation, and incorporated demand projections aligned to expected utilisation trends.

Each scenario was then evaluated against the baseline across five core dimensions: total cost of service delivery, cost per patient and per visit, GP workload and clinical time utilisation, patient access and service availability, and overall practice financial sustainability. All modelling assumptions were validated using a combination of current industry benchmarks from comparable GP practices, historical utilisation patterns and cost trends in primary care, and reference to prior modelling studies and published health system analyses. The model was developed with input from internal Discovery Health subject-matter experts through routine technical discussions and review processes. These engagements were conducted in the course of normal professional duties and were intended to inform the analysis and interpretation of findings, rather than to generate primary research data from human participants.

Sampling (case studies)

A purposive sampling methodology was used to select real-world private sector, nurse-led multidisciplinary primary health care (PHC) models. In the context of this study, purposive sampling ensured that the selected case studies reflect established and operational models of care that provide meaningful insight into how nurse-led PHC functions in practice and contributes to improvements in access, efficiency, and provider

incentives. The selection process focused on including a range of private sector and public–private partnership PHC models that demonstrate value-based care and innovation in PHC delivery within South Africa. Cases were drawn to capture variation across organisational forms, funding approaches, and service delivery platforms, while maintaining a consistent emphasis on multidisciplinary, nurse-led care. This enabled the study to examine how similar underlying principles, such as skill-mix optimisation, cost efficiency, coordinated care pathways and differentiated pricing structures, are implemented across different contexts. As a result, the case studies collectively provide a robust and contextually relevant evidence base to support the study's theoretical modelling framework and policy analysis.

Systematic review of literature (inclusion and exclusion criteria)

A structured systematic review of the literature was undertaken to inform the conceptual and empirical foundations of the study. The review focused on peer-reviewed journal articles, global health reports, and policy publications examining nurse-led care, MTBC, PHC delivery models, and provider payment mechanisms, with particular emphasis on LMIC contexts and comparators from high-income systems. Sources were identified through targeted searches of academic databases and institutional repositories, prioritising systematic reviews, randomised controlled trials, quasi-experimental studies, and authoritative global policy frameworks (e.g., WHO, OECD, and Lancet Commission publications).

- Inclusion criteria required studies to (i) evaluate nurse-led, task-shifting, or multidisciplinary team-based care models; (ii) report outcomes relating to at least one of the study's core domains (cost efficiency, access and service capacity, quality of care, or provider incentives); and (iii) demonstrate relevance to primary health care systems or first-contact care settings. Studies were included where they provided empirical evidence, robust modelling, or policy-relevant insights applicable to health system reform and workforce optimisation.
- Exclusion criteria applied to studies focused exclusively on tertiary or highly specialised care, those lacking measurable or comparable outcomes, and studies conducted in contexts with limited generalisability to PHC system design (e.g., highly fragmented or atypical delivery environments). Narrative opinion pieces without analytical grounding and studies with insufficient methodological transparency were also excluded. The final body of literature was synthesised thematically to identify consistent patterns in workforce allocation, clinical effectiveness, and system-level efficiency.

Sensitivity analysis

Given the nature of the hypothetical construct, various scenarios were tested to investigate the sensitivity of the results to variations in the key assumption on remuneration (fee per consultation) levels.

RESULTS

Team-based PHC delivery model

The team-based delivery model provides an illustrative comparison between a traditional solo GP practice structure and an integrated GP-and-nurse clinical team. In the baseline scenario, the standard general practitioner operates independently, seeing all patients and managing the full spectrum of clinical work. Under this model, the GP conducts an average of 4.2 visits per patient per year, spending approximately 15 minutes per consultation on average (based on observations from medical scheme data). Allowing for a 12% adjustment for leave and non-clinical time, the practice can realistically care for approximately 1,500 unique patients per annum. With a consultation fee of R550, the baseline configuration generates roughly R3.5 million in annual gross income from consultations (noting that the model does not account for any other procedure fees or dispensing fees). This conventional setup serves as the reference point against which the team-based model is evaluated.

Table 1: Calculation of baseline scenario patients

Factor	Value	Notes
Average visits per patient per annum	4.2	Based on medical scheme utilisation
Average time per consultation (minutes)	15	4 consultations per consulting hour
Average consultation hours per month	131	Takes account of average working days per month and 7 consulting hours per working day
Patients per year	1,487 (rounded to 1,500)	(28 consults per day × 223 working days) ÷ 4.2 consults per patient per annum
Gross consultation revenue at a fee of R550 per consultation	R3.5 million	131 consulting hours × 4 consults per hour × R550 × 12 months

The revised delivery approach assumes lower patient utilisation rates (3.5 visits per person per annum), a substantially lower cost per visit, and a structured clinical workflow in which nurses take on a major portion of consultations. The model applies a 36% reduction in cost per visit, bringing the fee to R350 (an illustrative amount that does not constitute a costing or proposal), but introduces nurse-led triage and referral pathways to maintain efficiency. In this setup, 40% of patients continue to see the GP, while the remaining 60% are fully managed by the nurse.

Evidence from primary care nurse-triage models shows variation in the proportion of visits escalated to a GP. For example, a Welsh case study of two nurse triage systems, presented as a conference poster, found that approximately 34% to 46% of triaged cases resulted in same-day or routine GP appointments (Elliott et al., 2019). While the protocol for the ESTEEM multicentre trial (Campbell et al., 2013) notes that approximately 50% of triage requests may be managed by telephone consultation without immediate escalation, international evidence indicates that nurse-led triage services in primary care frequently resolve a proportion of patient contacts independently, while referring the balance for GP assessment. Published studies and reviews show substantial variation between programmes, but referral or escalation rates commonly fall within a broad range of approximately 30% to 70%, reflecting differences in patient acuity, triage protocols, scope of nursing practice and organisational risk tolerance (Moll van Charante et al., 2006; Park et al., 2025). The sensitivity of the results to varying the referral rate is set out below.

In the model, nurse-only consultations are assumed to be 15 minutes long on average. Visits (for the more complex patients referred to the higher level of care) are assumed to include an average 10-minute nurse assessment followed by an average 15-minute focused GP consultation, resulting in an average 25-minute combined encounter. The nurse contributes meaningfully to upstream assessment and history taking, allowing the GP to concentrate on diagnosis, decision making, and the management of complex cases. To ensure quality and clinical governance, the GP allocates 15% of their time to supervising the nursing team. Nurse employment costs are set at R50,000 per month per employee (with a total cost to practice of R75,000 once overheads and operational expenses are included). Based on these parameters, the appropriate workforce structure equates to roughly 2.66 nurses per GP.

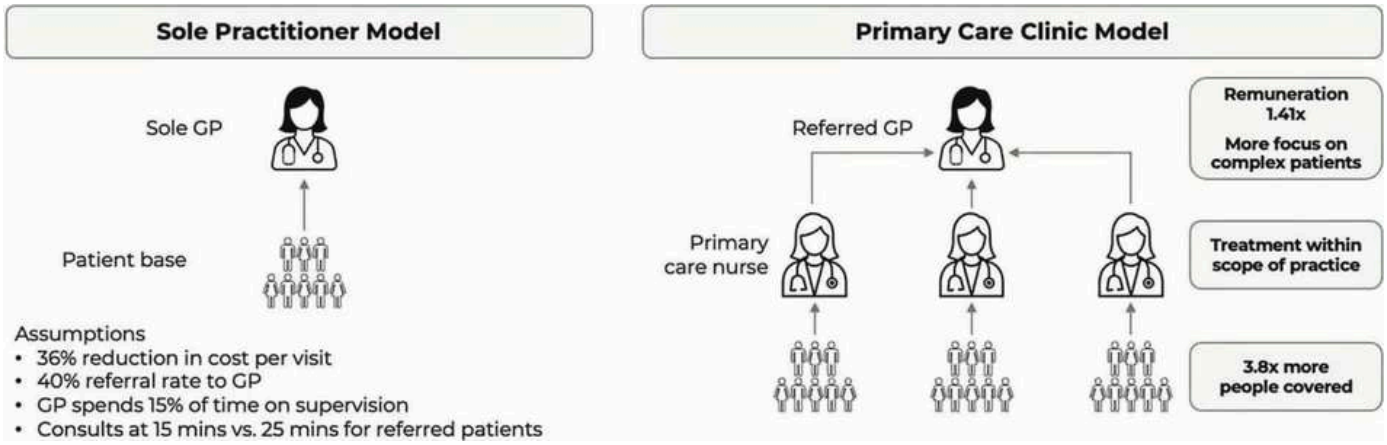

Figure 2: Schematic comparison of consultation services in the team-based model

This redesigned configuration significantly expands the practice's service capacity. With nurses absorbing most of the routine care, and with structured triage improving clinical flow, the team can care for approximately 5,700 unique patients per year (2.66 nurses each providing an assumed 33.6 visits per working day over 223 working days a year, at 3.5 visits per person per annum). This is about 3.8 times the number of patients achievable under the solo GP model. With 40% of about 89 team visits a day referred, this implies about 36 GP consultations a day, more GP consulting time than the 7 hours a day assumed for the solo GP. Even after deducting the costs of the additional nursing staff and practice overheads, net practice revenue to the GP (consultation revenue less nurse costs, not GP earnings) increases by about 41%, calculated on a monthly basis of 21 working days without the 12% leave adjustment. On these assumptions, a team-based arrangement could be both operationally feasible and financially advantageous for the clinician.

Table 2 below demonstrates the sensitivity to increasing the proportion of consultations that require the GP. This reduces the staff ratio as well as the additional revenue. A higher referral rate would therefore require an adjustment to the 36% discount of the consultation fee for the baseline model. This would raise efficiency questions relating to the clinical practice, although referral rates of 50% and 60% lie within the range reported in the literature cited above.

Table 2: Sensitivity testing for referral rate

Parameter	Single GP scenario	Model scenario	Scenario 2	Scenario 3
Consults with GP referral	100%	40%	50%	60%
Nurse ratio	0	2.66	2.04	1.63
Patient coverage	1,487	5,687	4,549	3,791
Patient coverage multiple	–	3.82	3.06	2.55
Additional GP net revenue	–	41%	15%	–3%

We note that the team-based PHC approach significantly enhances efficiency, improves financial performance, and allows GPs to specialise in more complex clinical decision-making, while increasing the total volume of care delivered (quality is not modelled directly). The model has specifically allowed for supervisory time to ensure (and potentially enhance) quality standards in the delivery of care.

Allocative efficiency and skill-mix optimisation in primary health care delivery

In the baseline configuration, the GP operates as a single, central resource responsible for all patient interactions. While this model reflects conventional primary care delivery structures, it introduces a binding

capacity constraint in which highly skilled GP time is allocated across both routine and complex clinical care activities. Evidence from advanced economies indicates that reliance on GP-led primary healthcare represents an inefficient allocation of health workforce resources, with significant opportunity costs to the health system. Specifically, when GPs are required to manage both routine and complex care, scarce and highly specialised clinical time is diverted away from higher-value clinically complex cases. This misallocation constrains overall health system capacity and reduces the number of patients that can be effectively served. In contrast, reallocating routine and lower-complexity care to nurses enables a more efficient use of available resources, thereby expanding access without proportionally increasing costs.

In support of the above, a substantial body of evidence from randomised controlled trials and systematic reviews of the literature demonstrates that nurse practitioner–led PHC delivery achieves clinical outcomes, patient satisfaction, and service utilisation patterns that are equivalent to those of physician-led care within their scope of practice, particularly for routine and chronic conditions (Martínez-González et al., 2014; Mundinger et al., 2000; Stanik-Hutt et al., 2013). Importantly, these models are associated with improved cost-effectiveness as they deliver comparable or better health outcomes at a lower cost per consultation, driven by differences in salary structures and consultation intensity. From an allocative efficiency perspective, nurse-led PHC models better align provider skills with patient needs by ensuring that each cadre operates within their scope of practice with the maximum level of operational efficiency and impact. This reduces inefficiencies associated with overqualification where highly trained GPs perform tasks that could be safely and effectively delivered by lower-cost providers, thereby optimising the overall distribution of health system resources. The resulting reconfiguration of care pathways not only enhances productivity but also improves system sustainability by enabling a shift from a high-cost, capacity-constrained model to a lower-cost, high-throughput model of care. The solo GP model thus allocates GP time to tasks that could be more efficiently delivered by nurses or allied health professionals. This results in a structural inefficiency where the primary constraint is not demand but the limited availability of GP time. The assumptions of 4.2 visits per patient per year and a 15-minute consultation duration reflect real-world utilisation; however, they also highlight how the exclusive reliance on GP-led care constrains throughput and restricts access.

In contrast, it has been argued that advanced practice nurses can deliver a large share of the primary care services traditionally provided by physicians, particularly in preventive care, chronic disease management, and health promotion (Nashwan, 2025).

This inefficiency becomes more pronounced when viewed against international workforce benchmarks and WHO recommendations on optimal skill mix. The World Health Organization emphasises that achieving UHC requires not only increasing workforce density but also optimising the distribution of tasks across cadres. Globally, nurse and midwife density (approximately 38.6 per 10,000 population) significantly exceeds physician density (approximately 16.7 per 10,000), reflecting the central role of nurses in delivering primary care services (WHO, 2022). In high-performing health systems, this imbalance is strategically leveraged through higher nurse-to-physician ratios, which average approximately 2.5:1 across OECD countries and exceed 4:1 in some systems, enabling physicians to focus on high-complexity care while nurses manage routine demand (Organisation for Economic Co-operation and Development [OECD], 2025). In contrast, health systems with lower nurse-to-physician ratios often exhibit inefficiencies, as physicians are required to perform tasks that could be safely delegated. This misalignment not only increases costs but also contributes to provider burnout without improving clinical outcomes.

Evidence from comparative primary care studies further indicates that nurse-led consultations often involve longer patient interactions, greater emphasis on patient education, and improved communication, which are associated with better adherence and patient satisfaction (Seale et al., 2005). Taken together, these findings suggest that the solo GP model represents a suboptimal configuration of health workforce resources. While clinically effective, it fails to leverage the full productive capacity of the broader primary care workforce. In contrast, nurse-led models with structured GP referral pathways are more consistent with WHO

recommendations, international workforce benchmarks, and the growing body of evidence supporting task shifting as a central mechanism for improving efficiency, expanding access, and enhancing the overall performance of primary healthcare systems.

Interpretation of economic and policy implications

From an economic perspective, the model provides important insights into how healthcare systems can be restructured to improve value. The shift from a high-cost, low-volume model to a lower-cost, high-volume model reflects a fundamental change in how services are delivered and financed. This has direct implications for policy design, particularly in the context of UHC and National Health Insurance. The ability to reduce costs while expanding access supports the sustainability of PHC funding models, making it feasible to extend coverage to a larger population. The model also highlights the importance of pricing as a policy tool. By setting consultation fees at a level that reflects underlying cost structures, the system can incentivise appropriate utilisation of services and encourage the adoption of efficient delivery models. In addition, the findings suggest that investment in workforce development, particularly for nurses, will be critical in supporting the transition to team-based care. This includes not only increasing the number of trained nurses but also expanding their scope of practice and providing appropriate training and support. Overall, the findings of this study provide a comprehensive view of how team-based PHC delivery can transform healthcare systems. The model demonstrates that significant gains in access, efficiency, and provider performance can be achieved through relatively straightforward changes in system design. The integration of modelling and case-study evidence supports the plausibility of these findings, although the proposed approach has not yet been tested in practice. The consistency between model outcomes and real-world case studies suggests that the transition to team-based care is not only possible but already underway. Importantly, the findings highlight that the benefits of the model extend beyond healthcare delivery. By improving access, reducing costs, and supporting workforce development, the model contributes to broader socioeconomic objectives, including equity, employment, and economic productivity.

In conclusion, the team-based PHC delivery model represents a potentially high-impact intervention that addresses multiple challenges within the health system simultaneously. Its scalability, affordability, and alignment with existing system dynamics make it a compelling strategy for strengthening primary healthcare and advancing broader health system reform.

Comparable case studies: a case for public–private partnerships (PPPs)

The National Development Plan (NDP) 2030 (National Planning Commission, 2012) recognises public–private partnerships (PPPs) as a strategic tool to improve health care funding, service delivery and infrastructure development in South Africa. According to the NDP, the role of PPPs in health includes:

- Emphasis on collaboration between the public and private sectors to address systemic challenges in healthcare.
- Leveraging private sector expertise, innovation, and financing to strengthen public health infrastructure and service delivery.
- Contributing towards equitable access to quality healthcare for all South Africans. Pillar 4 of the Presidential Health Compact (Republic of South Africa, 2019) also notes the importance of public–private partnerships in health care financing and service delivery.

Among other things, Pillar 4 advocates effective engagement of the private sector in improving health services access, coverage and quality. The Compact further states that "…over the past 25 years, different financing methods have been used to unlock access through building new public health infrastructure…. Some funding has been channeled through the National Department of Health and some from provincial equitable shares through provincial health departments…."

The WHO, the Organisation for Economic Co-operation and Development (OECD), and the World Economic Forum (WEF) have all highlighted the role of the private sector in achieving equitable, affordable access for citizens across the globe and recognised the potential value of PPPs in healthcare in delivering healthcare services efficiently and effectively (Hamdulay, 2023).

Table 3: Public–private partnership models in primary health care

Case study / region	PPP model	Government role	Key features	Positive outcomes	Source
COVID-19 PHC response	Multisectoral collaboration	Coordination	PHC teams with private doctors, nurses, CHWs and digital coordinators	Effective outreach, surveillance and vaccination outcomes	WHO case study
Gauteng: family medicine integration	Mixed staffing in PHC (private sector collaboration)	Collaboration with academic institutions to scale up family medicine training; promotion of continuous professional development and clinical governance	Family physicians, private GPs, public–private collaboration	Identified gaps and opportunities for PHC strengthening	African Journal of Primary Health Care & Family Medicine
Witkoppen Clinic, Gauteng	Hybrid PHC funding model	Trusted partner of the Gauteng Department of Health for over 25 years, collaborating on nurse training for antiretroviral treatment (NIMART), mentoring of pharmacy assistants and Expanded Public Works Programme cadres. These partnerships illustrate government's role in capacity building and skills transfer to strengthen PHC delivery in underserved areas.	Non-profit multidisciplinary model clinic offering comprehensive PHC; charges modest fees (around R150 for adults, R100 for children and pensioners); supported by donors, the private sector and government	Serves around 5,000+ patients a month; fills a gap for the uninsured; integrates social services and chronic care; scalable model for NHI inclusion	witkoppen.org.za; developmentaid.org
National HealthCare Group	Private managed care; employer-funded PHC	Private sector PHC provider that engages continuously with the NDoH on private PHC expansion and UHC	Offers PHC packages to low-income workers	Potential for employer–NHI partnerships	reachalliance.org
National: sessional GP contracting	Part-time GP services in public clinics	NDoH contracting and funding under NHI	GPs work sessionally in NHI pilot districts	Expanded access; reduced waiting times; improved care quality	gov.za
Clicks and Dis-Chem clinics	Integrated PHC retail model	Clicks and Dis-Chem[a] have multiple partnerships with the National Department of Health, most notably to make vaccines and contraceptives more accessible to public healthcare patients. Clicks clinics administer these services, with the government supplying the necessary stock	Clinics staffed by nurses, telemedicine access to GPs, PHC insurance	Affordable care for uninsured employed populations	Competition Commission
Unjani Clinics	Nurse-led social franchise model	Unjani Clinics do not operate as formal Department of Health facilities, but they complement public healthcare by relieving pressure on state clinics. Their model targets patients who can pay a small fee, reducing queues and resource strain in government facilities	Mostly black female professional nurses operate the clinics; supported by the private sector	Expanded PHC access in underserved areas; empowerment of female health entrepreneurs	Unjani Clinics website; academic case studies
Gauteng: STI care improvement	CME-based clinical quality improvement	District health managers were involved in planning and coordinating the study. The intervention explicitly sought to link private GPs with public health systems through surveillance and referral pathways	64 private GPs trained through CME seminars	Improved drug treatment for urethral discharge; potential for GP integration into NHI	Schneider et al. (2005)
GP contracting	Pilot contracting of private GPs	The National Department of Health led the GP Contracting Pilot by setting the contracting framework and compliance standards, ensuring alignment with NHI goals[b]	Performance-based contracts; integration with public health	Strengthened PHC delivery; scalable PPP model	Case study summary (Government of South Africa, 2012–2017)

[a] Dis-Chem clinics provide family planning (contraceptive injections and pills) and vaccinations in partnership with the NDoH, and collaborate with government on public health campaigns (e.g., COVID-19 testing, vaccination drives and family planning).

[b] The Department established performance indicators, monitored data for scalability, and supported GP training on public health protocols and electronic systems. It also coordinated with provincial health departments and private GP associations to integrate services into the District Health Information System (DHIS).

Note: CHW = community health worker; NDoH = National Department of Health; NHI = National Health Insurance; NIMART = nurse-initiated management of antiretroviral therapy; PHC = primary health care; PPP = public–private partnership; UHC = universal health coverage

DISCUSSION

Current access to primary healthcare in South Africa

Low-income households unable to afford traditional medical scheme cover have experienced consistently high levels of out-of-pocket expenditure on private primary healthcare, which is a regressive form of healthcare funding, particularly for this market. For example, the 2018 General Household Survey indicates that 27.1% of households use private facilities as their first point of contact for primary healthcare (Statistics South Africa, 2019); data from the NIDS study (Edwards, 2016) showed that 41.6% of the population last accessed care in the private sector; and recent experience from the exempted primary care insurance products reiterates the value that this market places on primary care services (it is estimated that 1.5 million to 2 million people have purchased exempted cover) (Moonstone Information Refinery, 2024). Primary health care (PHC) cover drives improved health outcomes, economic productivity and financial protection. Increasing privately funded cover for PHC has the potential to cover up to 11 million people in South Africa who live in households with employed members but who currently do not have medical scheme cover.

Expanding PHC cover is also aligned with WHO UHC principles. Public sector PHC funding has been declining over time, implying that patients are either paying out of pocket for private primary health care or delaying access. Further, the COVID-19 pandemic disrupted funding of, and access to, essential services in South Africa. A rapid assessment conducted by the WHO in 2020 found that 75% of countries reported disruptions to NCD services (South African Health Review, 2020). Budget allocations within the public sector had to be reprioritised in South Africa (National Treasury, 2020). Funding for several health care services within the PHC basket suffered, including HIV management (Jarolimova et al., 2021). The WHO recommendation is to reorient health systems to primary health care (PHC). It is estimated that most (90%) of essential UHC interventions can be delivered through PHC and that 75% of projected health gains from the SDGs could be achieved through PHC (WHO, 2024). This provides a compelling basis for investment in expanding access to primary healthcare services.

Strategic rationale for unlocking team-based PHC delivery models

The introduction of primary health care packages through Low-Cost Benefit Options (LCBOs) offers an opportunity to expand access to healthcare in a financially sustainable manner, thereby improving healthcare availability for all citizens, efficiency and cost-effectiveness. These primary health care products are also aligned to UHC objectives and can serve as a foundation for the National Health Insurance (NHI) benefit package. Primary health care products create opportunities for doctors to treat a larger patient base and contribute to improved health outcomes nationally. As noted above, these products could potentially reach up to 11 million people. At least 30% of South Africans currently access private primary care providers on an out-of-pocket (OOP) basis. This expenditure can be unpredictable, and costs vary widely for cash patients. The structure of these products will be simplified across medical schemes, enabling healthcare providers to unlock access to patients.

Proposed benefit package

The proposed team-based PHC delivery model is in line with the delivery of essential care and aligns with health priorities. It also includes unlimited access to care driven by healthcare need:

- The primary care packages are designed to address South Africa's most pressing health priorities, while ensuring affordability.
- Benefits include prevention and early detection, chronic disease management, maternal and child health, and HIV/AIDS care.

- Cover includes unlimited access to nurses with GP referrals as well as acute medication that enables population health management at a primary care level. The benefit package includes tests and screening for diagnosis and monitoring of key diseases.
- Cover can include comprehensive antenatal care including unlimited GP and nurse visits, essential blood tests, and ultrasound scans. These benefits support safer pregnancies and reduce maternal and infant mortality.
- Using healthcare resources appropriately through referral pathways means that there can be fair remuneration for services provided at all levels (i.e., nurses, GPs and pharmacists) because their skills are used appropriately.

	Unlimited primary care consults	Nurse consults with virtual and face to face GP referrals
	Acute medicine formulary	Built from the Essential Medicine List (EML)
	Chronic cover	For specific conditions such as Asthma, HIV, Diabetes, Hyperlipidemia, Hypertension etc.
	Antenatal and post-natal care	Included in primary care consults
	Mental health	Counselling and support (digital tools)
	Radiology and pathology	Formulary based on treatment protocols and referral pathways
	Preventive cover	Essential wellness screening, flu vaccine
	Optometry	Eye test and glasses formulary
	Dentistry	Basic dentistry formulary and preventive care

Figure 3: Illustrative package of primary care benefits

The cost of this benefit package is in the order of R400 per adult covered (as per the CMS LCBO Advisory Committee report; Council for Medical Schemes, 2023). However, for individuals earning above the 2025/26 tax threshold of R95,750 a year (approximately R7,979 a month; South African Revenue Service, 2025), cover would have minimal impact on take-home pay due to the medical scheme tax credit. The tax credit for 2025/26 was R364 for the main member and first dependent, and R246 for every subsequent dependant. In many cases, the tax credit nearly covers the full contribution implying minimal impact on take-home pay. For example, for a family of four with one working mother, one grandmother, and two children, the total monthly contribution would be R1,400. However, this employee would receive a tax credit of R1,220, implying that the net effect on her salary would only be R180 per month to cover her entire family. This is illustrated in Table 4 below.

Table 4: Illustrative cost of cover

Beneficiary category	Monthly contribution	Less tax credit*	Net monthly cost for taxpayers
Main member	R400	(R364)	R36
Adult dependant	R400	(R364)	R36
Child dependant	R300	(R246)	R54
Total for 2 adults and 1 child	**R1,100**	**(R974)**	**R126**
Total for 2 adults and 2 children	**R1,400**	**(R1,220)**	**R180**
Total for single parent and child	**R700**	**(R728)**	**Nil**

*Based on the current medical scheme tax credits applying to these options and the 2025/26 tax credits (South African Revenue Service, 2025).

Thought leadership: contribution to knowledge

This study advances the evidence base on primary health care (PHC) reform by integrating theoretically grounded modelling with empirical insights drawn from a diverse set of South African public–private partnership (PPP) case studies. In line with Siriram and Harris (2024), who highlight the need for structured and context-specific health economic modelling, this paper demonstrates how scenario-based modelling can be strengthened through alignment with real-world delivery innovations. The inclusion of case studies such as the COVID-19 PHC response, Witkoppen Clinic, Unjani Clinics, retail clinic networks, and GP contracting initiatives provides an empirical anchor that enhances the policy relevance and applicability of the modelling results.

Across the case studies, a consistent pattern emerges: PHC delivery in South Africa is already functionally nurse-led, reinforcing the structural observations made by Michel et al. (2018). For example, the COVID-19 PHC response demonstrated the effectiveness of multidisciplinary teams, anchored by nurses and community health workers (CHWs) and supported by digital coordinators, in delivering large-scale outreach, surveillance, and vaccination programmes. Similarly, retail clinic models such as Clicks and Dis-Chem rely predominantly on nurse-led service delivery, complemented by telemedicine access to GPs. These models illustrate how nurse-led care can operate at scale, particularly when supported by technology and structured referral pathways, confirming the central role of nurses in frontline PHC delivery across both public and private sectors. The Unjani Clinics model provides a particularly strong illustration of the intersection between service delivery, access expansion, and workforce development. As a nurse-led social franchise, Unjani Clinics empower predominantly black female professional nurses to operate independent practices while serving underserved populations at affordable price points. This aligns directly with Meintjies and Maritz (2025), who position nurse-led social entrepreneurship as a scalable and transformative career pathway.

The model not only expands access to care but also contributes to economic empowerment and workforce diversification, validating the dual service delivery and socioeconomic value of nurse-led PHC models. Case studies such as Witkoppen Clinic and the Gauteng Family Medicine Integration initiative further illustrate how multidisciplinary, team-based care can be operationalised while maintaining quality. Witkoppen Clinic, through its hybrid funding model and long-standing partnership with the Gauteng Department of Health, integrates nurse-led HIV care (including NIMART training), chronic disease management, and social services, serving over 5,000 patients per month. Similarly, the Gauteng Family Medicine Integration model highlights the importance of collaboration between private GPs, family physicians, and academic institutions in strengthening clinical governance and continuous professional development. These examples align with Fairall et al. (2012), who emphasise that nurse-enabled models can maintain or improve quality of care and patient safety when supported by structured supervision, training, and clinical governance frameworks. The GP contracting and sessional GP models provide further evidence of how team-based care can enhance

access and system efficiency. By integrating private GPs into the public sector through performance-based contracts, these initiatives have reduced waiting times, improved service availability, and strengthened PHC delivery at the district level.

Similarly, the Gauteng STI care improvement programme demonstrates how capacity building through continuous medical education (CME), combined with strengthened referral pathways, can improve clinical outcomes and integrate private providers into public health systems. These models reinforce the importance of coordinated, system-level approaches to PHC delivery. From a financing and policy perspective, several case studies illustrate how pricing and funding arrangements can be used to shape system performance, consistent with the framework proposed by Barber et al. (2018).

Employer-funded PHC models, such as those implemented by the National HealthCare Group, demonstrate how prepaid, lower-cost PHC packages can expand access to care among low-income workers while reducing reliance on out-of-pocket expenditure. Similarly, retail clinics and Unjani Clinics operate at price points that are significantly lower than traditional GP consultations, effectively using pricing as a mechanism to improve affordability and shift utilisation towards primary care. These observations directly support the modelling findings, which show that differential pricing combined with nurse-led triage can reduce average costs while maintaining provider incentives. Finally, the clinical and economic validity of nurse-led care is reinforced by parallels with the findings of Barton et al. (2013), who demonstrated that nurse-led antiretroviral treatment in South Africa is as cost-effective as doctor-led care. This is reflected in practice across several case studies, particularly in HIV and chronic disease management at facilities such as Witkoppen Clinic and in nurse-led retail and community-based clinics. These real-world examples provide strong empirical validation for the modelled outcomes presented in this study.

Collectively, the integration of these case studies with theoretical modelling allows this research to move beyond conceptual analysis and offer a deeply grounded, policy-relevant contribution to the literature. The convergence between empirical evidence and modelled outcomes demonstrates that nurse-led, team-based PHC delivery is not only theoretically robust but already partially realised within the South African health system. This positions the study as a significant contribution to thought leadership, providing a coherent and scalable framework for leveraging existing innovations to improve access, efficiency, and sustainability in pursuit of UHC and NHI objectives.

CONCLUSION

This study demonstrates that strengthening the primary health care (PHC) system requires a decisive shift from traditional, solo general practitioner (GP) practice models towards more efficient, cost-effective, and scalable team-based delivery approaches. While global evidence from the World Health Organization and the Lancet Commission has consistently highlighted the benefits of multidisciplinary, nurse-enabled care, this paper advances the discourse by applying context-specific, theoretically grounded modelling to the South African setting. In doing so, it directly responds to the need identified by Siriram and Harris (2024) for more robust, locally relevant health economic modelling frameworks to inform policy and system design. The findings reinforce that nurse-led, team-based care is inherently aligned with the realities of health systems in sub-Saharan Africa, where primary care delivery is already predominantly nurse-driven (Michel et al., 2018).

This study builds on that foundation by quantifying the extent to which structured task-sharing and nurse-led triage can expand capacity, reduce costs, and improve provider productivity. Importantly, the model allows GP supervision time so that these gains need not compromise quality of care, consistent with evidence showing that nurse-enabled models can maintain or enhance patient safety outcomes when supported by appropriate governance and supervision structures (Fairall et al., 2012). In the South African context, characterised by declining public sector funding, increasing demand for services and persistent inequities in access, the need for scalable delivery models is particularly urgent. The modelling suggests that

a nurse-led, team-based PHC approach can significantly increase system capacity and efficiency, while improving financial sustainability for providers.

This aligns with empirical evidence demonstrating the cost-effectiveness of nurse-led care relative to doctor-led care in the management of chronic conditions (Barton et al., 2013), reinforcing the economic rationale for task shifting as a core reform lever. From a policy and financing perspective, the study highlights how pricing structures and benefit design can be used strategically to influence provider behaviour and optimise system performance. The integration of differential pricing, nurse-led triage, and structured referral pathways provides a practical mechanism to lower the average cost of care while maintaining provider incentives, in line with the policy insights advanced by Barber et al. (2018). This underscores the importance of aligning payment systems with delivery model innovations to achieve sustainable health system reform.

The analysis further situates the proposed model within South Africa's evolving PHC ecosystem, where public–private partnerships (PPPs) and innovative delivery approaches are already demonstrating impact. Examples such as Witkoppen Clinic, Unjani Clinics, and employer-funded primary care networks illustrate the feasibility and scalability of blended service delivery models. These approaches also reinforce the growing importance of nurse-led service provision, including within emerging pathways of nurse-led social entrepreneurship (Meintjies & Maritz, 2025), which expand both access to care and professional opportunities within the health workforce. Scaling such models through low-cost benefit options (LCBOs) presents a significant opportunity to extend PHC coverage to millions of South Africans currently reliant on out-of-pocket spending. This would not only enhance financial protection and support UHC objectives but also generate substantial employment opportunities for both nurses and GPs.

Overall, this study contributes to the evidence base by modelling how nurse-led, team-based PHC delivery could be operationally feasible, economically advantageous, and contextually appropriate. It provides a credible and policy-relevant pathway for strengthening primary care and advancing South Africa's transition towards a more equitable, efficient, and sustainable health system under National Health Insurance (NHI).

REFERENCE LIST

Barber, S. L., Kumar, A., Roubal, T., Colombo, F., & Lorenzoni, L. (2018). Harnessing the private health sector by using prices as a policy instrument: Lessons learned from South Africa. *Health Policy*, *122*(5), 558–564. https://doi.org/10.1016/j.healthpol.2018.03.018

Barton, G. R., Fairall, L., Bachmann, M. O., Uebel, K., Timmerman, V., Lombard, C., & Zwarenstein, M. (2013). Cost-effectiveness of nurse-led versus doctor-led antiretroviral treatment in South Africa: Pragmatic cluster randomised trial. *Tropical Medicine & International Health*, *18*(6), 769–777. https://doi.org/10.1111/tmi.12093

Campbell, J. L., Britten, N., Green, C., Holt, T. A., Lattimer, V., Richards, S. H., Richards, D. A., Salisbury, C., Taylor, R. S., & Fletcher, E. (2013). The effectiveness and cost-effectiveness of telephone triage of patients requesting same day consultations in general practice: Study protocol for a cluster randomised controlled trial comparing nurse-led and GP-led management systems (ESTEEM). *Trials*, *14*, Article 4. https://doi.org/10.1186/1745-6215-14-4

Council for Medical Schemes. (2023). *Low-cost benefit options (LCBO): Recommendations and guidelines report*.

Edwards, T. (2016). *Socio-demographic determinants of health-seeking behaviour among the South African population: An analysis of NIDS* [Master's dissertation, University of KwaZulu-Natal]. ResearchSpace. https://researchspace.ukzn.ac.za

Elliott, M., Jones, S., Johnson, C., & Wallace, C. (2019). *A geographical case study of two primary care nursing triage systems in Wales* [Poster presentation]. Rural Health and Care Conference. https://pure.southwales.ac.uk/en/publications/a-geographical-case-study-of-two-primary-care-nursing-triage-syst/

Fairall, L., Bachmann, M. O., Lombard, C., Timmerman, V., Uebel, K., Zwarenstein, M., Boulle, A., Georgeu, D., Colvin, C. J., Lewin, S., Faris, G., Cornick, R., Draper, B., Tshabalala, M., Kotze, E., van Vuuren, C., Steyn, D., Chapman, R., & Bateman, E. (2012). Task shifting of antiretroviral treatment from doctors to primary-care nurses in South Africa (STRETCH): A pragmatic, parallel, cluster-randomised trial. *The Lancet*, *380*(9845), 889–898. https://doi.org/10.1016/S0140-6736(12)60730-2

Floriancic, N., Garnett, A., & Donelle, L. (2024). Chronic disease management in a nurse practitioner-led clinic: An interpretive description study. *SAGE Open Nursing*, *10*. https://doi.org/10.1177/23779608241299292

Frikha, Y., Freeman, A. R., Côté, N., Charette, C., & Desfossés, M. (2024). Transformation of primary care settings implementing a co-located team-based care model: A scoping review. *BMC Health Services Research*, *24*, Article 890. https://doi.org/10.1186/s12913-

24-11291-7

Government of South Africa. (2012–2017). *GP contracting pilot case study/summary under NHI*. https://www.gov.za/

Hamdulay, A. (2023, September 14). Partnerships are critical in building a sustainable healthcare system in South Africa. *FAnews*. https://www.fanews.co.za/article/healthcare/6/general/1124/partnerships-are-critical-in-building-a-sustainable-healthcare-system-in-south-africa/37919

Hanson, K., Brikci, N., Erlangga, D., Alebachew, A., De Allegri, M., Balabanova, D., Blecher, M., Cashin, C., Esperato, A., Hipgrave, D., Kalisa, I., Kurowski, C., Meng, Q., Morgan, D., Mtei, G., Nolte, E., Onoka, C., Powell-Jackson, T., Roland, M., . . . Wurie, H. (2022). The Lancet Global Health Commission on financing primary health care: Putting people at the centre. *The Lancet Global Health*, *10*(5), e715–e772. https://doi.org/10.1016/S2214-109X(22)00005-5

Jarolimova, J., Yan, J., Govere, S., Bunda, B. A., Ngobese, N., Shazi, Z. M., Khumalo, A. R., Nelson, G., Ngcobo, L. M., Wara, N. J., Zionts, D., Parker, R. A., & Bassett, I. V. (2021). Decentralised HIV care during the COVID-19 pandemic: Ensuring safe community-based services. *South African Medical Journal*, *111*(8), 698. https://doi.org/10.7196/SAMJ.2021.v111i8.15864

Kim, D. D. (2024). The role of simulation modelling in public health policy evaluation [Comment]. *The Lancet Public Health*, *9*(3), e150–e151. https://doi.org/10.1016/S2468-2667(24)00027-6

Kruk, M. E., Gage, A. D., Joseph, N. T., Danaei, G., García-Saisó, S., & Salomon, J. A. (2018). Mortality due to low-quality health systems in the universal health coverage era: A systematic analysis of amenable deaths in 137 countries. *The Lancet*, *392*(10160), 2203–2212. https://doi.org/10.1016/S0140-6736(18)31668-4

Larrain, N., & Groene, O. (2021). Simulation modeling to assess performance of integrated healthcare systems: Literature review to characterize the field and visual aid to guide model selection. *PLOS ONE*, *16*(7), Article e0254334. https://doi.org/10.1371/journal.pone.0254334

Liang, Z., Montgomery, J., Dingelstad, C., Koschel, A., & Redford, M. (2025). Making multidisciplinary team-based care (MTBC) a success in primary care: A scoping review. *Journal of Healthcare Leadership*, *17*, 681–695. https://doi.org/10.2147/JHL.S560629

Martínez-González, N. A., Tandjung, R., Djalali, S., Huber-Geismann, F., Markun, S., & Rosemann, T. (2014). Effects of physician–nurse substitution on clinical parameters: A systematic review and meta-analysis. *PLOS ONE*, *9*(2), Article e89181. https://doi.org/10.1371/journal.pone.0089181

Mathauer, I., & Wittenbecher, F. (2013). Hospital payment systems based on diagnosis-related groups: Experiences in low- and middle-income countries. *Bulletin of the World Health Organization*, *91*(10), 746–756A. https://doi.org/10.2471/BLT.12.115931

Meintjies, C. S., & Maritz, J. E. (2025). Nurse-led social entrepreneurship as a career. *Health SA Gesondheid*, *30*, Article a2700. https://doi.org/10.4102/hsag.v30i0.2700

Michel, J., Evans, D., Tediosi, F., de Savigny, D., Egger, M., Bärnighausen, T., McIntyre, D., & Rispel, L. (2018). Lest we forget, primary health care in sub-Saharan Africa is nurse led. Is this reflected in the current health systems strengthening undertakings and initiatives? *Journal of Global Health Reports*, *2*, Article e2018009. https://doi.org/10.29392/joghr.2.e2018009

Moll van Charante, E. P., ter Riet, G., Drost, S., van der Linden, L., Klazinga, N. S., & Bindels, P. J. E. (2006). Nurse telephone triage in out-of-hours GP practice: Determinants of independent advice and return consultation. *BMC Family Practice*, *7*, Article 74. https://doi.org/10.1186/1471-2296-7-74

Moonstone Information Refinery. (2024). *CMS extends exemption for primary healthcare policies*. https://www.moonstone.co.za/cms-extends-exemption-for-primary-healthcare-policies/

Mundinger, M. O., Kane, R. L., Lenz, E. R., Totten, A. M., Tsai, W. Y., Cleary, P. D., Friedewald, W. T., Siu, A. L., & Shelanski, M. L. (2000). Primary care outcomes in patients treated by nurse practitioners or physicians: A randomized trial. *JAMA*, *283*(1), 59–68. https://doi.org/10.1001/jama.283.1.59

Nashwan, A. J. (2025). Transforming primary care in low- and middle-income countries through advanced practice nursing. *The Lancet Primary Care*, *1*(3), Article 100033. https://doi.org/10.1016/j.lanprc.2025.100033

National Planning Commission. (2012). *National Development Plan 2030: Our future – make it work*. The Presidency, Republic of South Africa. https://www.gov.za/sites/default/files/gcis_document/201409/ndp-2030-our-future-make-it-workr.pdf

National Treasury. (2020). *Supplementary budget review 2020*. https://archive.treasury.gov.za/documents/national%20budget/2020S/review/FullSBR.pdf

Ogunbekun, I., Ogunbekun, A., & Orobaton, N. (1999). Private health care in Nigeria: Walking the tightrope. *Health Policy and Planning*, *14*(2), 174–181. https://doi.org/10.1093/heapol/14.2.174

Organisation for Economic Co-operation and Development. (2025). *Health at a glance 2025: OECD indicators*. OECD Publishing. https://doi.org/10.1787/8f9e3f98-en

Park, H., Shin, D.-S., Wade, C., & Duong, D. B.-L. (2025). Scoping review of nurse triage in primary care. *BMC Nursing*, *24*, Article 1104. https://doi.org/10.1186/s12912-025-03740-3

Reach Alliance. (2023). *Social franchising as a model to improve access to health care in South Africa*. University of Toronto. https://reachalliance.org/reports/social-franchising-as-a-model-to-improve-access-to-health-care-in-south-africa/

Republic of South Africa. (2019). *Presidential Health Compact*. The Presidency. https://www.presidency.gov.za/sites/default/files/2022-05/Presidential%20Health%20Compact.pdf

Schneider, H., Chabikuli, N., Blaauw, D., Funani, I., & Brugha, R. (2005). Improving the quality of STI care by private general practitioners: A South African case study. *Sexually Transmitted Infections*, *81*(5), 419–420. https://doi.org/10.1136/sti.2004.010470

Seale, C., Anderson, E., & Kinnersley, P. (2005). Comparison of GP and nurse practitioner consultations: An observational study. *British Journal of General Practice*, *55*(521), 938–943.

Siriram, C., & Harris, R. (2024). The need for health economic model guidelines in South Africa. *South African Journal of Economic and Management Sciences*, *27*(1), Article a5798. https://doi.org/10.4102/sajems.v27i1.5798

Sitthimongkol, Y., Srimoragot, M., Kasemsuk, W., Meekusol, S., Pongpattanapisit, P., Saenkla, P., & Ruksakulpiwat, S. (2025). Enhancing accessibility through nurse-led clinics in primary care: An integrative review of models of care. *International Journal of Nursing Sciences*, *12*(6), 593–600. https://doi.org/10.1016/j.ijnss.2025.10.006

South African Health Review. (2020). *COVID-19 service disruptions and NCDs*. Health Systems Trust.

South African Revenue Service. (2025). *Budget tax guide 2025/26*. https://www.sars.gov.za/wp-content/uploads/2025/06/Budget-tax-guide-21-May-2025.pdf

Stanik-Hutt, J., Newhouse, R. P., White, K. M., Johantgen, M., Bass, E. B., Zangaro, G., Wilson, R. F., Fountain, L., Steinwachs, D. M., Heindel, L., & Weiner, J. P. (2013). The quality and effectiveness of care provided by nurse practitioners. *The Journal for Nurse Practitioners*, *9*(8), 492–500. https://doi.org/10.1016/j.nurpra.2013.07.004

Statistics South Africa. (2019). *General household survey 2018*.

Terry, D., Hills, D., Bradley, C., & Govan, L. (2024). Nurse-led clinics in primary health care: A scoping review of contemporary definitions, implementation enablers and barriers and their health impact. *Journal of Clinical Nursing*, *33*(5), 1724–1738. https://doi.org/10.1111/jocn.17003

Togoveanu, R. (2023). The use of simulation modeling in cost-effectiveness analysis of pharmacoeconomics [Opinion]. *Pharmacoeconomics: Open Access*, *8*(3), 178. https://doi.org/10.37421/2472-1042.2023.8.178

Unjani Clinics NPC. (n.d.). *About the Unjani Clinics network*.

Van Weel, C., & Kidd, M. R. (2018). Why strengthening primary health care is essential to achieving universal health coverage. *CMAJ*, *190*(15), E463–E466. https://doi.org/10.1503/cmaj.170784

Witkoppen Health and Welfare Centre. (n.d.). *Witkoppen Health and Welfare Centre* [Website]. https://www.witkoppen.org/

World Health Organization. (2019a). *Declaration of Astana: Global Conference on Primary Health Care, Astana, Kazakhstan, 25 and 26 October 2018* (WHO/HIS/SDS/2018.61). https://iris.who.int/handle/10665/328123

World Health Organization. (2019b). *Primary health care on the road to universal health coverage: 2019 monitoring report*.

World Health Organization. (2021). *Global expenditure on health: Public spending on the rise?* https://www.who.int/publications/i/item/9789240041219

World Health Organization. (2022). *Global health workforce statistics database* [Data set].

World Health Organization. (2024). *Universal health coverage: Report by the Director-General* (EB156/6). https://apps.who.int/gb/ebwha/pdf_files/EB156/B156_6-en.pdf

World Health Organization & United Nations Children's Fund. (2018). *A vision for primary health care in the 21st century: Towards universal health coverage and the Sustainable Development Goals*. https://www.who.int/publications/i/item/WHO-HIS-SDS-2018

DECLARATIONS

Acknowledgements

The authors wish to acknowledge the valuable support provided by colleagues at Discovery Health in the preparation and proofreading of this paper. The authors also acknowledge that they are employees of Discovery Health (Pty) Ltd, an accredited medical scheme administrator.

Author contributions

Roseanne Harris: Conceptualisation, methodology, supervision, writing – review & editing;

Nondumiso Gugu Beatrice Khumalo: Literature survey, case study curation, writing – original draft;

Sarika Besesar: Model review, writing – review & editing.

All authors have read and approved the final version of this manuscript.

Data availability statement

The data supporting the findings of this study are available from the authors upon reasonable request, subject to ethical approval and institutional regulations.

Declaration of competing interest

All authors are employees of Discovery Health (Pty) Ltd, an accredited medical scheme administrator, and the model presented in the paper is built on depersonalised medical scheme data. The authors have been involved in preparing regulatory inputs and exemption applications relating to LCBOs.

Declaration of ethics

The authors declare that the research and writing of this manuscript adhere to ethical standards of research and publication, in accordance with scientific principles, and are free from plagiarism. The study collected no primary data from human participants: the model uses depersonalised medical scheme data, and Discovery Health subject-matter experts contributed in the course of their normal professional duties.

Declaration of assistive technologies in the writing process

The authors declare that generative artificial intelligence (GenAI) and AI-assisted tools were used in a limited and responsible manner during the preparation of this manuscript. The authors take full responsibility for the final content of this manuscript.



STREAM II
HTA AND PRIORITY SETTING

TOWARD UNIVERSAL HEALTH COVERAGE: REIMAGINING THE HEALTH TECHNOLOGY ASSESSMENT FRAMEWORK IN SOUTH AFRICA

Mark A. Brand

How to cite this article: Brand, M. A. (2026). Toward universal health coverage: Reimagining the health technology assessment framework in South Africa. *Council for Medical Schemes Research Journal*, *1*(1), 59–77.

ABSTRACT

Background: Health technology assessment (HTA) in South Africa is described as underdeveloped and fragmented, primarily focused on reimbursement and cost containment rather than on promoting patient access. Processes are characterised by secrecy and variable lead times, are medicines-centric and rely heavily on randomised controlled trials, while failing to account for the unique characteristics of medical technologies such as shorter product lifecycles and iterative improvements. This paper advocates for the establishment of a single, independent, transparent and accountable HTA agency for South Africa, underpinned by empowering legislation aligned with progress towards universal health coverage (UHC), and tailored to accommodate the distinctive characteristics of medical technologies. The paper focuses specifically on processes and methodologies designed for medical devices, equipment and in vitro diagnostics (IVDs), collectively defined as medical technology (MedTech).

Methods: International HTA models and their relevance to the evolving South African policy landscape are analysed. The paper explores the relationship of regulator, HTA body and coverage bodies in an ecosystem in which these are independent of, but interdependent with, each other, and recommends inclusion of the assessment of therapeutic and economic value followed by appraisal by independent, multistakeholder committees. The analysis explores models of governance, structure, processes and methods of assessment for varying health technologies to recommend a sustainable, transparent, independent and accountable HTA framework. Challenges with applying medicines-focused HTA methods to MedTech are discussed and bespoke processes and methods of assessment proposed.

Results: This paper proposes a lean, government-funded coordinating HTA public office, supported by independent, multistakeholder appraisal committees including clinicians, patient groups, relevant technology experts and industry representatives to ensure objectivity. Evaluation pathways should be developed for each technology, such as relative effectiveness assessments, and real-world data should be used to identify and encourage the adoption of innovative technologies that integrate effectively into clinical pathways.

Conclusions: A locally relevant, legislatively empowered and credible HTA mechanism that safeguards independence, transparency, participation and accountability, and that leverages existing public and private capacity to support equitable priority setting and value-informed coverage decisions, is vital for South Africa's journey towards universal health coverage.

KEYWORDS

Health technology assessment; medical technology; National Health Insurance; universal health coverage; governance; reimbursement

INTRODUCTION

Health technology assessment (HTA) is a multidisciplinary process that uses explicit methods to determine the value of a health technology throughout its lifecycle. Its purpose is to inform decision-making to promote equitable, efficient, and high-quality health systems (O'Rourke et al., 2020). The World Health Organization (WHO) defines health technologies broadly as the application of organised knowledge and skills in the form of devices, medicines, vaccines, procedures and systems developed to solve a health problem and improve quality of life (World Health Organization, n.d.). Globally, HTA is a mature policy instrument for guiding resource allocation, improving patient access, and rationalising health system expenditure. In developed economies, HTA is relatively advanced, comprehensive, and well-resourced, with a primary focus on improving patient access to health technology. By contrast, in developing countries, including South Africa, HTA is fragmented and used as a cost-containment or cost-restriction mechanism, with comparatively less emphasis on patient access.

This divergence has direct implications for South Africa, an upper-middle-income country (World Bank Group, n.d.), which is navigating a transition towards universal health coverage (UHC) via the National Health Insurance (NHI) system.

Currently, local HTA practice lags significantly behind global standards. Activities across the public and private sectors are uncoordinated, insufficiently transparent, and lack independent oversight. Local exercises are confined to individual medical scheme clinical policies, formulary decision-making and cost-cutting. They lack formal topic selection, prioritisation mechanisms, or broader stakeholder consultation. This stands in stark contrast to established international agencies that maintain transparent review processes, independent appraisals by multisectoral representative committees, and mandatory public recommendations (Battista & Hodge, 2009; Castro, 2017; Towse et al., 2011).

The National Department of Health's Essential Drugs Programme took a step closer to formalisation of HTA by publishing the HTA Methods Guide (National Department of Health, 2023). However, the guide is limited to medicines on the Essential Medicines List, and stakeholder feedback on the draft noted that it lacked a definitive process and did not acknowledge the unique lifecycles or bespoke assessment needs of other health technologies. A critical knowledge and systemic gap in South Africa is the routine and inappropriate application of medicines-centric HTA methods to medical technologies (MedTech).

Medical technologies function as complex interventions and differ fundamentally from medicines across several core domains (Drummond et al., 2009; Tarricone et al., 2017a):

- Innovation cycles: MedTech is characterised by rapid, continuous, and iterative improvements. Traditional, lengthy HTA timelines mean scientific evidence can become outdated or irrelevant before the assessment concludes.
- Evidence generation: MedTech generally presents lower levels of classical scientific evidence (Schnell-Inderst et al., 2015). Conducting randomised controlled trials (RCTs) often poses severe ethical issues. Furthermore, clinical efficacy is heavily influenced by operator experience, clinical learning curves, and device upgrades.
- Commercial realities: Medical devices have significantly shorter product lifecycles and highly compressed investment recovery periods compared to medicines.

When South African assessment bodies rigidly apply pharmaceutical metrics to complex devices, innovative technologies fail their assessments, yielding predictably negative outcomes for patient care across prevention, screening, monitoring, diagnosis, treatment and rehabilitation. International surveys confirm that standardised, homogenised HTA procedures for MedTech remain uncommon, raising questions about whether a uniform approach should be implemented across distinct technology types (Ciani et al., 2015; Fuchs et al., 2016; Jacobs et al., 2017). There are specific aspects to consider in the methods for assessing medical technology, including topic selection, scoping and communication with manufacturers, timing of assessments, dynamic development of new technologies, and relatively low levels of evidence (Schnell-Inderst et al., 2015).

These distinctions demand bespoke assessment methodologies and dedicated governance frameworks. Multiple local policy frameworks, including the National Health Act (Act 61 of 2003; Republic of South Africa, 2004), the National Health Insurance Act (Act 20 of 2023; Republic of South Africa, 2024), the Health Market Inquiry (Competition Commission of South Africa, 2019) and the Presidential Health Compact (South African Government, 2019) explicitly call for institutionalised HTA, built on the principles of independence, transparency and stakeholder participation. However, the complete absence of empowering legislation serves as a major structural barrier. Without a formal legislative scaffold, any generated HTA recommendations lack legitimacy and legal enforceability. This legislative vacuum represents one of the most significant structural barriers to the development of a credible national HTA system.

This paper reflects on an institutionalised HTA mechanism, advocating for an HTA structure that is independent, transparent, and accountable to all its stakeholders and beneficiaries, understanding the particular characteristics of different health technologies rather than methods designed primarily for medicines (O'Rourke et al., 2020; Schnell-Inderst et al., 2015). Medical technology is presented as vital across the continuum of patient care, spanning prevention, screening, monitoring, diagnosis, treatment, and rehabilitation. The topic is therefore framed as a health system issue that affects access, utilisation, and value across multiple sites of care. An important distinction between regulatory, HTA and coverage (reimbursement) functions is emphasised. Similarly, HTA differs from guideline development, although the two are closely related, interdependent processes that are often used together to inform health policy, with the latter informed by HTA (Cheyne et al., 2023). Regulation of health technologies falls within the mandate of the South African Health Products Regulatory Authority (SAHPRA), whereas HTA activities, which inform reimbursement and resource allocation decisions, are currently undertaken by multiple entities.

It is argued that a system independent of the regulatory function enables decisions about safety to be separated from decisions about allocation of resources, thereby facilitating clearer and more accountable decision-making. This distinction becomes central to the proposed governance recommendations and to the preferred structural archetype for South Africa (Allen et al., 2013, 2017; Wild et al., 2017). The paper explores the current status of HTA internationally and locally, examines the specific challenges of HTA for medical technologies, analyses governance models and structural archetypes, proposes a concrete HTA structure and process framework for South Africa, and outlines a strategy for institutionalisation. In so doing, this paper seeks to inform the design of an HTA system capable of supporting South Africa's progress towards UHC by ensuring HTA becomes part of the decision-making process in the health system, even if initially limited in its use to specific innovative medical technologies.

LITERATURE REVIEW

HTA in developing versus developed contexts

The evolution of HTA as a global policy instrument is extensively documented (Battista & Hodge, 2009). In developed economies, HTA agencies are well-established, comprehensively resourced, and operate with a clear mandate to improve patient access to effective health technologies. In contrast, HTA conducted in

emerging markets, including South Africa, is fragmented and used as a cost-containment or cost-restriction mechanism rather than a value-identifying tool (Towse et al., 2011). Consequently, HTA appears not to be used in an effective way by these healthcare systems (Towse et al., 2011). This fundamental reorientation from value identification to cost restriction constitutes a significant gap in the application of HTA in developing economies, and one that an institutionalised South African HTA framework must consciously seek to reverse.

The natural history of HTA has demonstrated that its development follows a broadly predictable trajectory, from fragmented activity towards institutionalised, legislatively anchored practice (Battista & Hodge, 2009). Castro (2017) documented the policy process of establishing an HTA agency in Colombia, highlighting the critical role of political support, legislative frameworks, and stakeholder alignment in institutionalising HTA. The Colombian experience is instructive for South Africa, where similar policy-level ambition exists but lacks the legislative scaffolding required to translate intent into enforceable mechanisms. A South African Health Review analysis further noted that a unified vision for HTA in South Africa would require broad stakeholder engagement, policy alignment with the NHI, harmonisation of legislation, and ultimately parliamentary enactment (Siegfried et al., 2017).

Governance principles for HTA agencies

Three core governance values of independence, transparency, and accountability are found to exert major influence over the success of HTA organisations (Lafortune et al., 2008). Organisation of HTA requires the broadest possible scope in its appraisal of all evidence, and its processes must be deliberately designed to accommodate this scope (Drummond et al., 2008). These principles are supported by a WHO Global Survey of HTA Agencies, which provides insight into impediments and opportunities for developing appropriate HTA models in resource-constrained contexts. This survey concluded that developing a system to use HTA requires not only methodological capacity but also governance structures, including legislation with clear lines of accountability for decision-making, and that political support and leadership are essential (WHO, 2015, p. 29).

Developing partnerships with academic institutions was further identified as a particularly useful strategy for low-resource settings (WHO, 2015, p. 29). Wild et al. (2017) provided detailed guidance on the development of national HTA strategies, emphasising that an HTA structure achieving a segregation of functions is rooted in the HTA value of independence from interest groups and political interference of any kind. Common lessons from the challenges of HTA and its successful implementation have been extensively researched and reported (Drummond et al., 2008), and these lessons should be systematically considered in the design and development of HTA for South Africa. A well-functioning regulatory authority (in South Africa's case, SAHPRA) is also identified as critical to the success of HTA, as an HTA system that is independent of the regulatory function enables decisions about resource allocation to be separated from decisions about safety.

Structural archetypes for HTA systems

A systematic analysis of the pathways for regulatory (REG), health technology assessment (HTA) and coverage decisions (CB) across Europe, conducted using a mapping methodology to illustrate HTA organisation relationships and their integration with regulators and coverage bodies, identified two taxonomic groupings of national HTA systems and ten distinct HTA archetypes (Allen et al., 2013, 2017). The first taxonomic grouping (Figure 1) is based on the positional relationship between a national HTA agency (HTA), its regulatory body (REG), and the coverage body (CB), and whether these functions are combined or performed independently of one another.

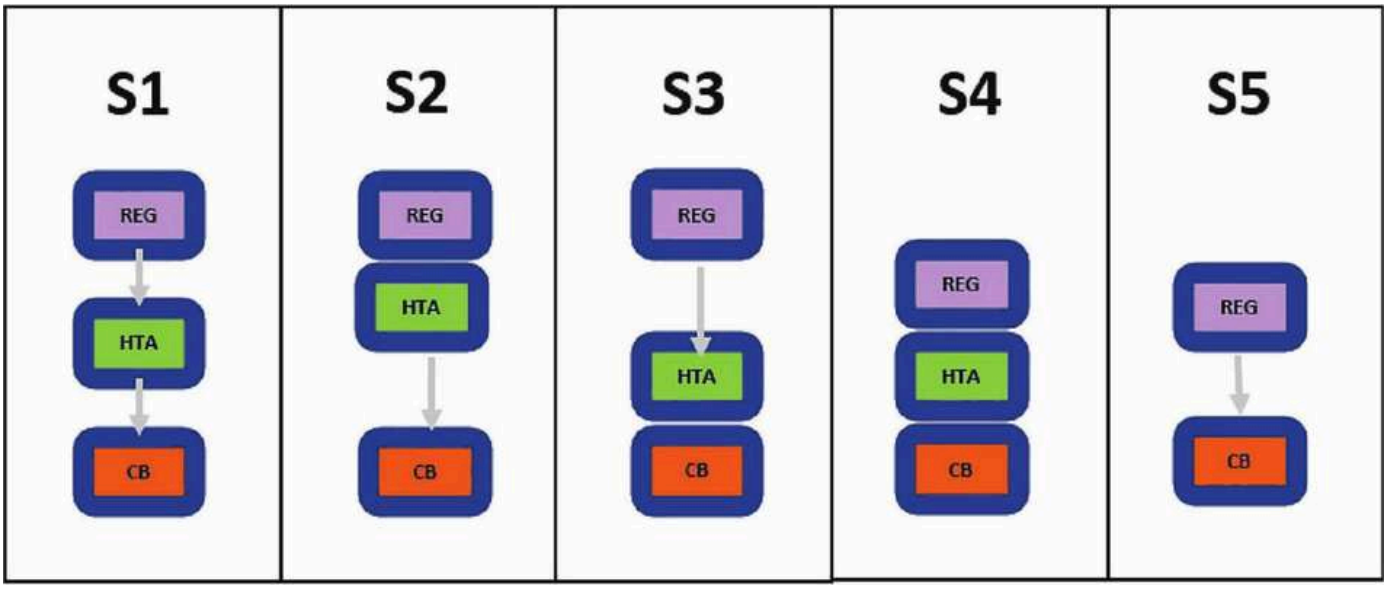

Figure 1: System process taxonomy (adapted from Allen et al., 2013)

Note: REG = regulator; HTA = health technology assessment; CB = coverage body. **S1:** Regulatory, HTA and coverage body functions are performed by separate agencies. **S2:** Regulatory and HTA functions are performed by a single agency, and the coverage body functions are independent. **S3:** HTA and coverage body functions are performed by a single agency with the regulatory function performed independently. **S4:** Regulatory, HTA and coverage body functions are all performed within a single agency. **S5:** No HTA is performed within the national regulatory to reimbursement system.

The second grouping (Figure 2) is based on the three key tasks performed by the HTA agency: therapeutic value (TV) assessment, economic value (EV) assessment, and appraisal (AP). The inclusion of an independent appraisal step can reduce the perception of bias and is therefore an important factor that should be considered for the purposes of classification (Drummond et al., 2008).

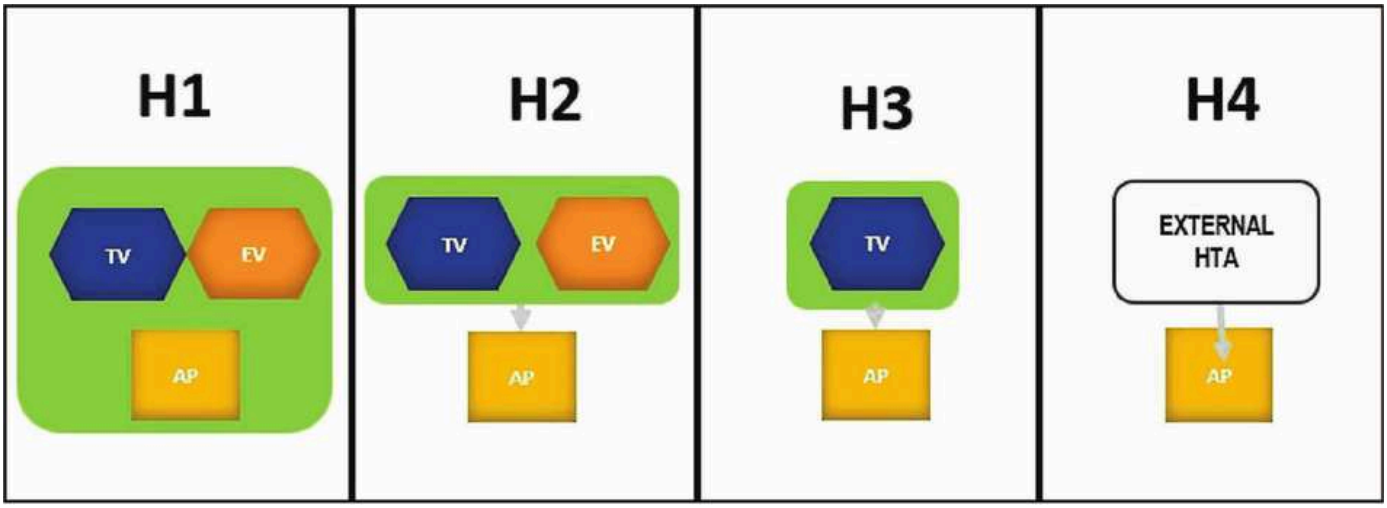

Figure 2: HTA process taxonomy (adapted from Allen et al., 2013)

H1: Therapeutic value assessment, economic evaluation and appraisal are performed within the same agency. **H2:** Therapeutic value assessment is conducted within the same agency as economic evaluation, but the appraisal is performed independently, usually by health professionals rather than civil servants. **H3:** Therapeutic value is assessed prior to independent appraisal. **H4:** Appraisal is conducted using information from an external HTA report or by considering the coverage decisions of reference countries.

In most European cases, HTA is separate from the regulator, and in about half of these systems the HTA agency also performs the appraisal function (Allen et al., 2013). The interactions between these three core functions (the regulator, HTA and coverage body) can ultimately affect overall system performance (Allen et al., 2013). The history of HTA internationally underscores that it requires a completely different set of skills

and types of evidence from those needed for regulatory review and appraisal, making it critical to segregate these functions (Banta & Jonsson, 2009).

Challenges in HTA of medical technologies

The academic literature consistently acknowledges that HTA methods for medicines are incompatible with medical technologies without significant and deliberate modification (Tarricone et al., 2017a). Medical technology should be conceptualised as complex interventions, requiring the use of appropriate approaches that recognise and accommodate their distinct characteristics (Tarricone et al., 2017a). Fuchs et al. (2016) conducted a comprehensive review of HTA of medical devices in Europe, finding significant heterogeneity in processes, practices, and methods across European national HTA systems. A concurrent survey of non-European Union HTA agencies similarly revealed that standardised procedures for medical technology from the perspective of HTA institutions are not common (Ciani et al., 2015), raising the fundamental question of whether a homogeneous approach should be implemented across all health technologies (Jacobs et al., 2017). Specific methodological challenges for MedTech HTA include the use of observational data to assess comparative effectiveness, estimating and anticipating learning curves, studying the diffusion and adoption of technologies, and coverage with evidence development (Tarricone et al., 2017b). Fuchs et al. (2017), examining HTA of medical devices from a European perspective, identified further structural, procedural, methodological, and regulatory challenges and proposed directions for future improvement. Daubner-Bendes et al. (2021) specifically examined HTA for medical devices in Central and Eastern Europe and recommended that in relatively late HTA adopter countries, medical technology suppliers should be able to leverage reported HTA recommendations from other countries where the technology is already in use. The availability of scientific evidence for medical technologies is substantially lower than for medicines, not least because of the ethical issues associated with conducting randomised controlled trials (RCTs) for devices, the continuous iterative changes and upgrades, learning curves and operator experience effects, and the challenges of selecting appropriate endpoints.

Many new and innovative medical technologies fall short in their South African assessments due to a perceived lack of evidence, while assessment bodies persist in applying medicines methodologies yielding predictably negative outcomes. HTA agencies seldom differentiate between methods and processes used for medicines and medical technologies (Ciani et al., 2015; Fuchs et al., 2016).

International models of best practice

Several international HTA agencies have developed explicit processes and methods guides for medical technology. Agencies such as the National Institute for Health and Care Excellence (NICE) in England, the Medical Services Advisory Committee (MSAC) in Australia, the Canadian Agency for Drugs and Technologies in Health (CADTH), and the Singapore Agency for Care Effectiveness (ACE) have all recognised the need for specific evaluation pathways for medical technology. These agencies design pathways to identify, evaluate, and encourage adoption of new and innovative medical technologies, including how the technology integrates into current clinical pathways and its comparative effectiveness relative to current best practice.

The European Network for Health Technology Assessment (EUnetHTA) Project was established in 2006 out of an informal initiative between a small number of European HTA agencies to create an effective and sustainable network for HTA across Europe. This has progressed into a more sustainable, permanent collaboration for health technology assessment (HTA) in Europe, evolving into joint HTA work between agencies with certain deliverables including Joint Scientific Consultations (JSC) and Joint Clinical Assessments (JCA), and to lay out the framework and methodological details of the European HTA procedures and groundwork for HTA Regulations. This EUnetHTA collaboration produced the Core Model 3.0 (European Network for Health Technology Assessment [EUnetHTA], 2016), known as a standardised, transparent framework for conducting and reporting health technology assessments in Europe, enabling

cross-border collaboration. While the EUnetHTA cooperation officially expired in September 2023, work of EU HTA bodies continues in accordance with the new EU HTA Regulations (HTAR).

This model provides a common structure, with nine HTA domains including health problem and current use of technology, description and technical characteristics, safety, clinical effectiveness, cost and economic evaluation, ethical analysis, organisational aspects, patient and social aspects, and legal aspects, with the latter four collectively referred to as ELSO domains (EUnetHTA, 2016). The model enables standardised reporting and provides a common framework for the production, presentation, and use of HTA information, and is identified as a relevant international framework for adoption or adaptation in the South African context. Rapid Relative Effectiveness Assessments (REA) within this framework are intended for assessments within a limited time frame, covering only the first four domains, and are identified as a pragmatic alternative pathway for many medical devices.

METHODOLOGY

The methodological approach is qualitative and is primarily characterised by document analysis, comparative policy review, and stakeholder-informed synthesis rather than primary empirical data collection. As a review paper, it does not report findings from an experimental or observational study but rather synthesises available evidence and expert positions to formulate recommendations for HTA institutionalisation in South Africa.

Research design

The research design is qualitative and exploratory in nature, employing a policy analysis framework that systematically examines existing HTA governance models, structural archetypes, legislative frameworks, and methodological approaches from both international and South African perspectives. The design is consistent with the nature of a scoping review paper, in which an evidence-informed normative argument is constructed from the systematic assembly and critical appraisal of existing sources. The paper is not designed for replication in the sense of an empirical study; rather, its replicability lies in the transparency of its source base and the logical coherence of its analytical and recommendation framework.

Sources of evidence and document analysis

The primary source base encompasses peer-reviewed journal literature, grey literature including policy documents and official reports, and international HTA agency guidance materials. South African policy documents reviewed include the National Health Act (2003) and preceding papers and Bill, NDoH Health Technology Strategy documents, NHI Act (2023), the Health Market Inquiry, the Presidential Health Summit Compact (South African Government, 2019), and HTA Methods Guide (National Department of Health, 2023). International bodies and frameworks referenced include the WHO, EUnetHTA, the International Network of Agencies for Health Technology Assessment (INAHTA), the UK National Institute for Health and Care Excellence (NICE), Australian Medical Services Advisory Committee (MSAC), the Canadian Agency for Drugs and Technologies in Health (CADTH), and the Singapore Agency for Care Effectiveness (ACE), among others. Peer-reviewed literature was drawn from international journals covering HTA methodology, health policy, and health economics, with a focus on sources directly relevant to the challenges of assessing medical technologies. Both recent and seminal sources were incorporated: foundational literature on the history and development of HTA (Banta & Jonsson, 2009; Battista & Hodge, 2009) was reviewed alongside current developments in MedTech-specific methodology (Daubner-Bendes et al., 2021; Fuchs et al., 2016, 2017; Tarricone et al., 2017a, 2017b).

Comparative analysis of HTA governance models

A structured comparative analysis was conducted of HTA governance models internationally, drawing on a mapping methodology originally applied to European HTA systems by Allen et al. (2013). The two taxonomic sets described above were cross-referenced to help identify practical archetype groupings of HTA agencies, with each group indicating an archetype relative to their position in the system, and the tasks they perform (see Figure 3). Ten distinct HTA archetypes were identified; in most cases, HTA is separate from the regulator, and in about half of these systems the HTA agency also performs the appraisal. The interactions between these three core functions (the regulator, HTA and the coverage body [payer]) can ultimately affect overall system performance.

This is then applied as a reference framework for identifying the most appropriate structural archetype for South Africa. The analysis specifically focused on the degree of independence between the regulatory function, the HTA function, and the coverage/reimbursement decision-making function, as this segregation is identified as a core determinant of HTA quality and legitimacy (Wild et al., 2017).

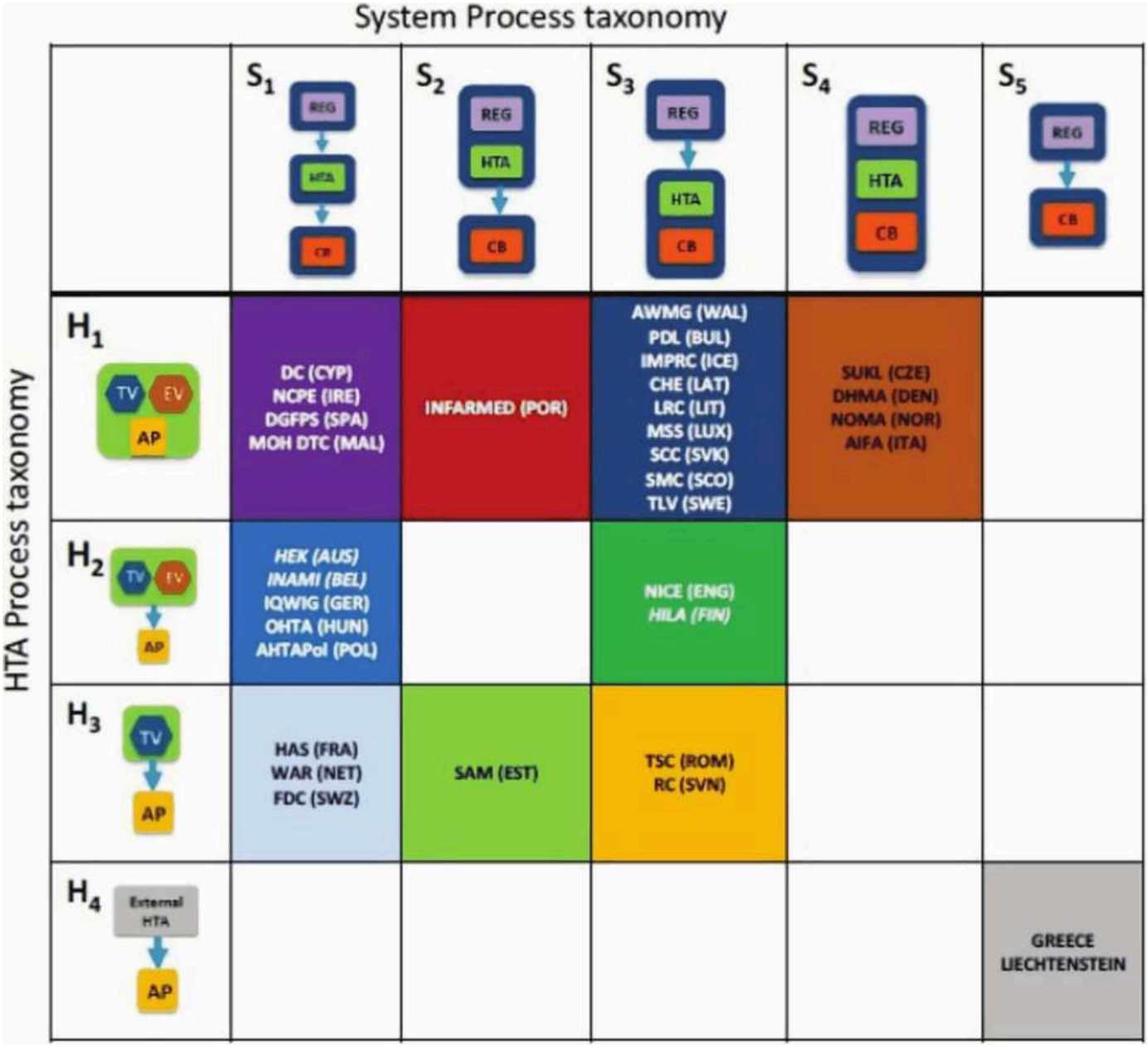

Figure 3: Practical archetype groupings of European HTA agencies

Note: This grid shows the allocation of each HTA agency to process taxonomy group and its respective national system. The ten archetype groups are listed by their acronym abbreviations in the archetype key (Figure 4). Empty boxes have no agencies. Adapted from Allen et al. (2013).

Process and methods design

The proposed HTA process framework for medical technology was developed by examining the traditional five-step HTA process used by existing agencies and adapting it to the specific characteristics of MedTech. Key process design features identified as important include: developing innovative approaches for assessment, estimating learning curves, using observational data to assess comparative effectiveness, studying the diffusion of technology, and coverage with evidence development (Tarricone et al., 2017b). The EUnetHTA Core Model was identified as a standardised methodological framework adaptable for South Africa (EUnetHTA, 2016), and the INAHTA's three HTA assessment strategies (Rapid Reviews, Mini HTAs and Full HTAs) were incorporated into the tiered assessment design (Malaysian Health Technology Assessment Section [MaHTAS], n.d.).

Table 1 summarises the HTA design principles on which the proposed HTA process (Figure 6, Results section) is based.

Table 1: Summary of HTA process design principles for medical technology

HTA phase	Key functions	Key principles	Medtech-specific features
Initiation and planning	Topic selection; prioritisation; routing; project scoping	Goal and scope explicit; new class or category only; any stakeholder may initiate	Use of GMDN classification; multistakeholder prioritisation committee
Assessment and analysis	Literature reviews; evidence synthesis; appointment of external evaluators	Independently conducted; wide range of evidence; include international HTA agencies	Observational data; learning curves; coverage with evidence development (CED)
Appraisal	Appraisal of assessment; gap analysis; therapeutic and economic value; multi-criteria decision analysis (MCDA)	Unbiased and transparent; different perspectives of value; multistakeholder committees	ELSO domains included; accountability for reasonableness (A4R)
Recommendation	Communication strategy; public database; pilot studies (CED); implementation plan	All outputs in public domain; explicit appeals process; transparent link to decision-making	Multistakeholder appeals committee; mandatory inclusion in benefits packages
Implementation	Optimisation; monitoring of adoption; new topic identification; training oversight	Site of care for optimum delivery; monitor implementation of findings	Communication of findings to all key decision-makers; credentialling oversight

Note: A4R = accountability for reasonableness; CED = coverage with evidence development; ELSO = ethical, legal, social and organisational; GMDN = Global Medical Device Nomenclature; HTA = health technology assessment; MCDA = multi-criteria decision analysis.

RESULTS

Current status of HTA for medical technology in South Africa

Analysis of the HTA landscape in South Africa reveals a system that is highly fragmented, variable in quality, and insufficiently aligned with international standards of independence and transparency. HTA activities are conducted by numerous actors across public and private sectors (including medical aid schemes, administrators, managed care companies, private hospitals, professional societies and academic institutions) but are uncoordinated, with no prioritisation or topic selection mechanism and limited stakeholder consultation. The scope of HTA activities performed trends towards assessment of safety, performance, quality and price referencing, which are regulatory rather than HTA aspects, irrespective of the international

and local regulatory status and real-world evidence. In the private funder sector, new medical technologies are assessed to inform reimbursement decisions, with medical schemes acting as primary gatekeepers for market access.

Submission templates differ across assessment bodies, limiting the quality and comparability of information provided. Systematic research is conducted internally by the assessing organisation, with comparative evidence, ideally RCTs, preferred, while grey and unpublished literature is excluded. Engagement with clinical stakeholders and industry is limited, and patient input is often not considered. Final recommendations may be made anywhere from 6 to 24 months after application, or may be delayed indefinitely, with in some cases no opportunity to appeal within 12 months of notification of the decision unless new evidence is available.

A negative outcome of assessment translates into no reimbursement, non-approval by hospitals and no access to formularies, and consequently patient access is denied. A medical scheme may decide to reimburse for richer plan options only, with no funding available on lower plans; alternatively, the scheme may reimburse but exclude the medical technology component if limits or formularies are in place. There is neither provision nor request for information of any ethical, legal, social, or organisational impact in such assessments; such information may be volunteered by an applicant but is not necessarily considered as relevant. These dynamics are contrary to international trends, where HTA agencies maintain transparent review processes, independent multisectoral appraisals, and mandatory publicly available recommendations (Castro, 2017).

Legal and policy framework analysis

Regulation of health technologies falls within the mandate of the South African Health Products Regulatory Authority (SAHPRA), whereas HTA activities are primarily applied to private sector reimbursement decisions by medical schemes. HTA in South Africa has been an evolving discussion for over two decades, featured in several policy documents as discussed previously. A common thread throughout these documents is the need for the establishment of institutionalised HTA, founded on the principles of independence, transparency, and participation; there is, however, an absence of empowering legislation for the formal establishment of an HTA agency, without which recommendations will lack legitimacy and therefore enforceability.

NHI envisages the establishment of HTA within the NHI as a transitional arrangement, with an interim Ministerial Advisory Committee on HTA (MAC-HTA) proposed to oversee the organisation of such an entity, reporting to the Minister of Health and advising the NHI Benefits Advisory Committee (BAC). Critically, the latest NHI Act is not explicit regarding either the independence or the permanency of an HTA agency within the NHI structures. The implication is that the Essential Drugs Programme (EDP) may become the core of an HTA Agency in South Africa, which is likely to experience the same challenges assessing medical technologies as has been the case in other countries, with subsequent difficulties implementing recommendations.

HTA governance models and structural archetypes

The comparative analysis of European HTA systems identified ten distinct HTA archetypes, which are mapped in Figure 3 and listed in Figure 4 (Allen et al., 2013).

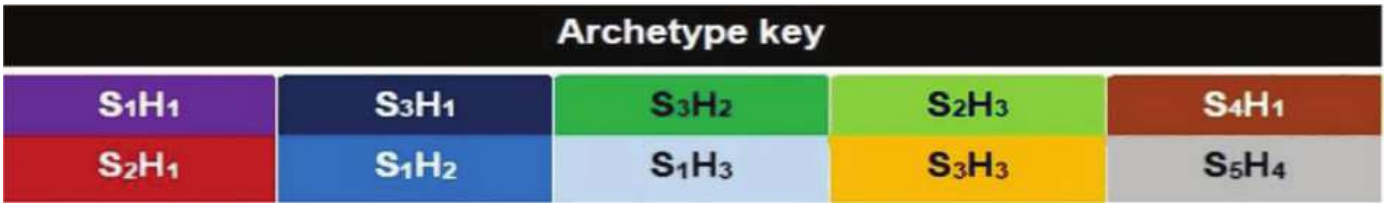

Figure 4: Mapped HTA archetypes across European countries

Note: Adapted from Allen et al. (2013).

In most cases, HTA is separate from the regulator, and in about half of these systems the HTA agency also performs the appraisal function (Allen et al., 2013). The interactions between the three core functions (regulator, HTA and coverage body) can ultimately affect overall system performance (Allen et al., 2013).

The S1H2 archetype emerges as the most appropriate structural model for an HTA system in South Africa. This archetype implies that the regulator is separate from the HTA Agency, the coverage body and final decision-maker, with therapeutic value and economic value assessments conducted within the same agency, and appraisal performed by an independent, multi-representative Appraisal Committee. A multistakeholder, multisectoral Interim Steering Committee for HTA (ISC-HTA), appointed by the Minister of Health in 2004 after studying HTA systems of different countries, previously made proposals regarding the scope and structure for HTA, guidelines for prioritisation of HTA activities, and the creation of an HTA Agency, considering three possible options for the legal status of such a body: a legal person, a coordinating/collaborative office of current activities, or expanding the mandate of an existing body.

Proposed HTA structure for South Africa

The proposed HTA structure for South Africa is illustrated in Figure 5. Based on international experience and the evolution of HTA agencies in other markets, there is a strong case for leveraging existing capacity and coordinating all such activities for a common purpose, primarily to reduce duplication of work, improve standardisation, reduce bias in appraisals, encourage transparency, increase credibility, and improve levels of compliance.

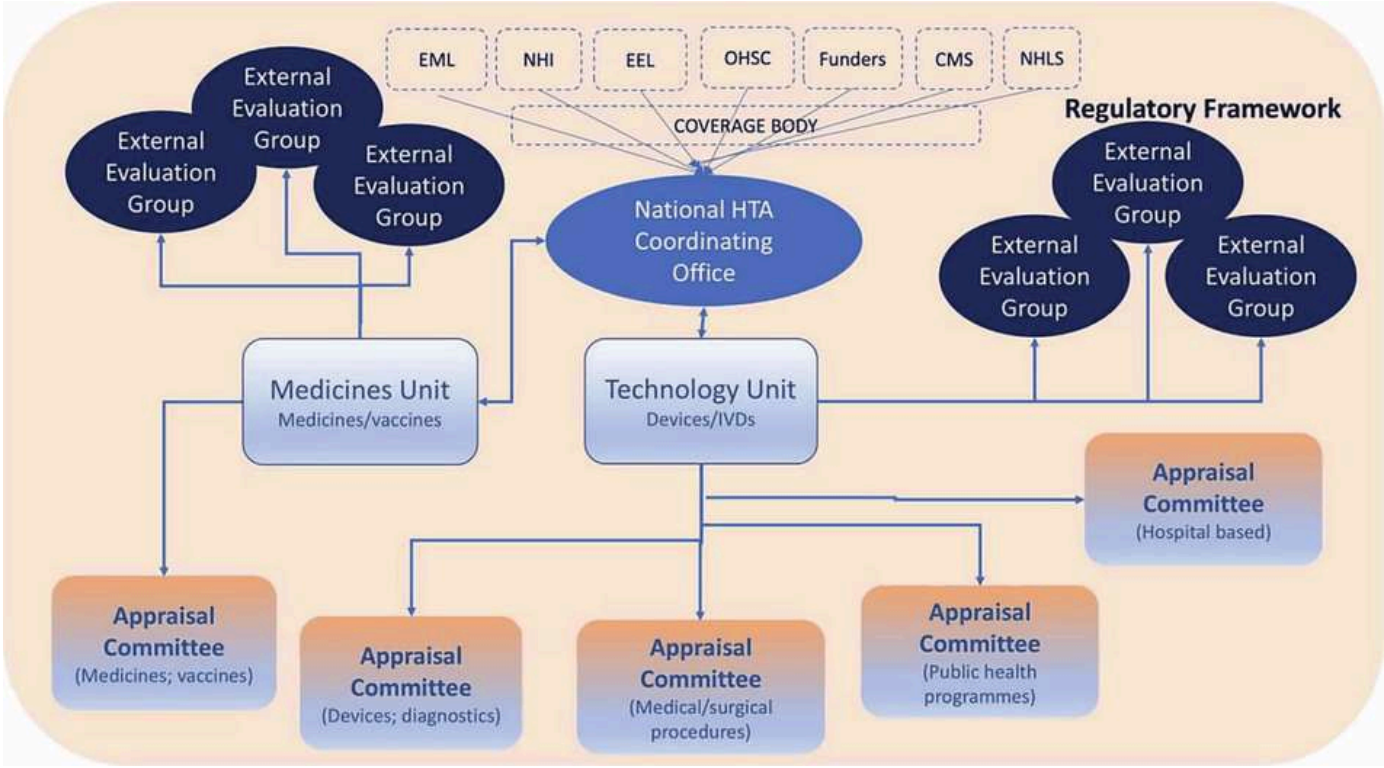

Figure 5: Health technology assessment ecosystem

Note 1: Examples of coverage bodies (i.e., users of HTA outcomes; not an exhaustive list). EML: Essential Medicines List; NHI: National Health Insurance; EEL: Essential Equipment List; OHSC: Office of Health Standards Compliance; CMS: Council for Medical Schemes; NHLS: National Health Laboratory Service. Coverage bodies may also be part of appraisal committees. Note 2: Examples of appraisal committees relevant to the technology under assessment, indicating who, and what experience, should be represented on each committee.

The proposed structure comprises six interrelated components:

1. A National HTA Coordinating Office serving as the entry point for coordinating HTAs requested by coverage/decision-making bodies or any stakeholder within the public or private sectors.

2. A Medicines HTA Unit responsible for HTA of health products including medicines and vaccines.

3. A Technology HTA Unit responsible for HTA of medical technologies including devices and IVDs. A third unit could address interventional procedures but is likely to be related and aligned to the medicines and technology units.

4. Contracted external academic evaluation groups providing analytical capacity and expertise.

5. Independent expert appraisal committees, each focusing on a different technical area, for example, medicines/vaccines, diagnostics and devices, medical/surgical procedures, public health programmes and healthcare institutions (operational requirements). Members are to be nominated from relevant stakeholder groups with the requisite expertise. Appraisal committees may be similar to the scoping committees, that is, have the same member expertise; and

6. An underlying regulatory framework articulating governance arrangements, appeal/review rights and processes, the composition and responsibilities of independent advisory committees, and the scope and limits of the Minister of Health's authority as final decision-maker.

Terms of Reference for each relevant group will need to be created. This model emulates Australian and Canadian models and is being adopted systematically in developing markets within the means of each

country's economy. This model suggests a hybrid approach within the context of limited but developing resources: a central body (likely under the National Department of Health [NDoH]) that coordinates HTA activities but heavily relies on external experts to do the heavy lifting.

Advantages include resource efficiency, by complementing limited government resources with external HTA specialists; smart matching, by allowing external researchers to focus on their specific areas of expertise; and strong oversight, with the office acting as a central air traffic controller for both internal and external evaluations.

Disadvantages may include perceived bias: stakeholders might feel the agency is too close to the NDoH and lacks true independence. Strict legislation and governance would be needed to address this, with the South African Medical Research Council offering a successful precedent.

Proposed HTA process for medical technology

The proposed HTA process for medical technologies follows a five-step approach, designed as a dynamic process with feedback loops providing for continuous review (see Figure 6). Common functions of each phase are listed, underpinned by relevant functions and major principles that need to be considered in process design. This proposed framework provides a non-exhaustive list of functions and principles per phase, to be unpacked and refined after stakeholder engagement in the development process. After market authorisation by SAHPRA, any stakeholder, including suppliers, may initiate a submission for HTA for a new medical technology; escalation to independent HTA is contingent only on the medical technology passing regulatory requirements, although it is not uncommon to conduct HTA in parallel to the regulatory process where there is sufficient recognition of therapeutic and economic value. This may be determined in the planning phase, which will follow separate processes.

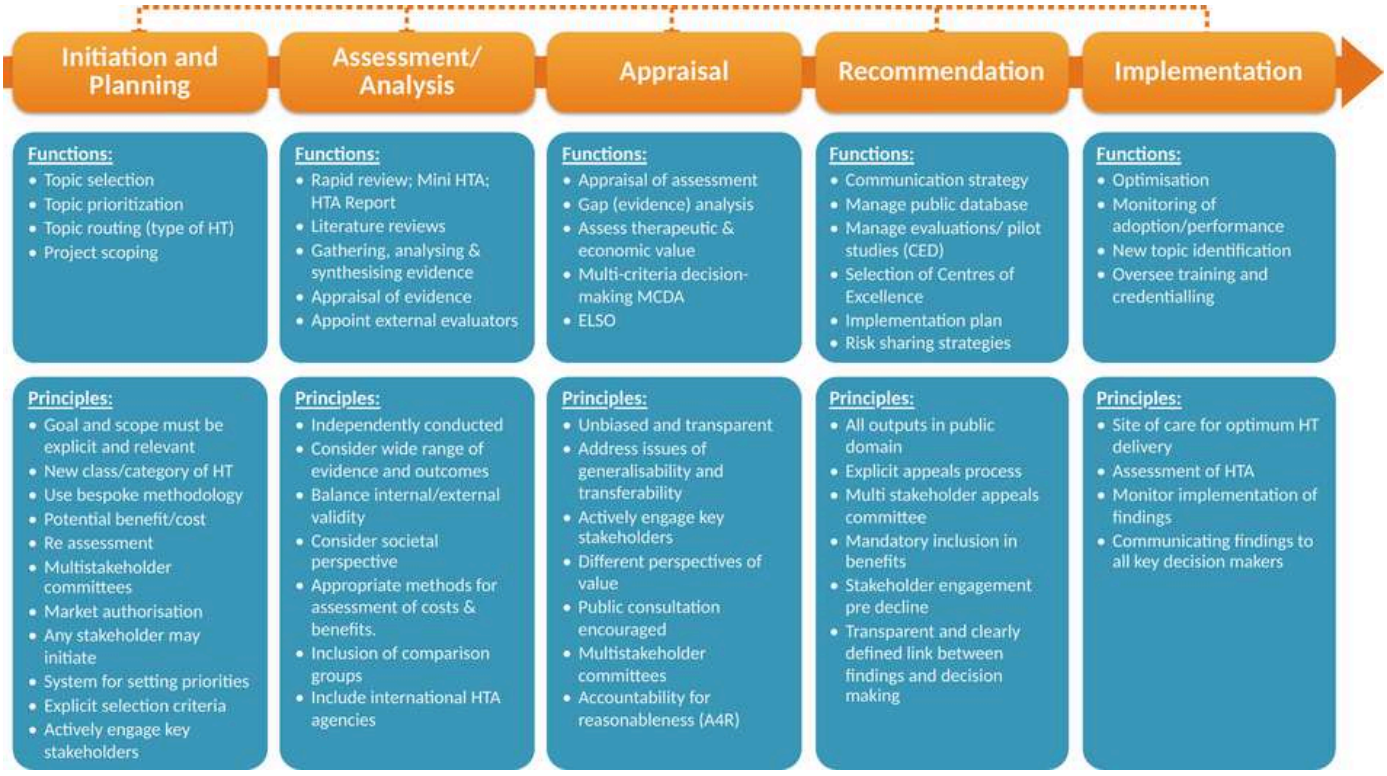

Figure 6: HTA process for medical technologies

Initiation and planning will include a multistakeholder committee, which will determine technology type, assessment pathway and scope. Technologies identified for HTA should be selected using explicit criteria. A

primary set of criteria may include innovation, relative risk (class), prioritisation, adoption, clinical need, and value statement.

Given the diversity and variable complexity of medical technologies, a three-tiered process is proposed for HTA routing to expedite recommendations and avoid unnecessary delays to patient access: Rapid Reviews, Mini HTAs, and Full HTAs, drawn from the INAHTA framework (MaHTAS, n.d.). The scope of these strategies must be fully understood and developed for the local context, the terms of reference and scope of devices defined for each, and relevant pathways developed, based on a multistakeholder process. The assessment and analysis phase will collate and synthesise all relevant evidence that befits a comprehensive HTA, and based on predetermined routing, will be referred to the relevant external evaluation groups. The synthesised assessment and analysis (including appraisal of synthesised evidence) is referred to the appraisal phase, where an external appraisal committee will consider therapeutic and economic value, including appraisal of all other HTA domains, including ELSO where applicable. A recommendation is based on all the evidence before it, after wide stakeholder consultation. It may determine a cost-effectiveness threshold but not recommend a price for that technology, as HTA should not assume the role of a pricing authority. The HTA Agency will formulate an implementation plan and communication strategy based on the relevant recommendations received.

Strategy framework for HTA institutionalisation

The institutionalisation of HTA is a multifaceted process that moves HTA from an ad hoc activity to a core, sustainable component of a health system's decision-making. It is not just about technical analysis; it is about creating a permanent ecosystem. This requires establishing accepted processes, norms and standards, developing critical skills, experience and knowledge, and building effective working relationships between stakeholders to support its use. Assessing medical technologies is highly complex and requires adaptability; a rigid, top-down, "one-size-fits-all" strategy will not work. The strategy must accommodate different technologies' unique operating contexts.

The medical technology community should be actively invited to help develop and implement these HTA processes.

The foundations for institutionalising HTA begin with an audit. Before building something new, evaluate existing HTA activities, identify necessary stakeholders, and assess current capacity. HTA may already exist under different names, such as "evidence-based medicine" or "cost-effectiveness analysis". A stakeholder analysis gathers information to see whose interests must be considered, who will support or block the initiative, and how to mitigate opposition. Involved groups include patients, policymakers, doctors, hospitals, academics, industry, donors, and the public. Consider "on-the-ground teams", in which clinical stakeholders (e.g., doctors) act as "champions" to solve institution-specific problems. An implementation team with the right skills and authority must be set up at the healthcare facility level to drive these multifaceted changes.

A significant barrier to HTA development is the skills gap, where successful institutionalisation requires developing technical capacity to both produce and interpret HTA reports and identifying the specific training needs of different groups to ensure the process is technically sound and inclusive. The vision for HTA in South Africa expressed in the *South African Health Review* (Siegfried et al., 2017) involves a unified vision aligning HTA with the NHI goals through legislative alignment, harmonising policies and passing necessary parliamentary amendments; consensus building through stakeholder engagement to find agreement on definitions and the potential for a national HTA agency; and a strategic framework, to adopt a model typical of low- and middle-income countries (LMICs) that focuses on political commitment and efficient information provision. This strategy must be underpinned by a commitment to establish a political framework to promote (enforce, facilitate) HTA uptake and establish structures for timely, efficient and good-quality provision of HTA information (Wild et al., 2017).

DISCUSSION

Interpreting findings in the context of the literature

The findings presented in this paper are broadly consistent with the international literature on HTA development in emerging economies. Towse et al. (2011) observed that HTA in emerging markets is predominantly used as a cost-containment tool rather than a value identification mechanism, and this pattern is clearly evident in South Africa, where HTA-like activities in the private sector focus primarily on cost benchmarking rather than comprehensive value assessment.

The misapplication of medicines-focused methodologies to medical technologies further exacerbates this problem, as documented extensively in the European context by Fuchs et al. (2016) and Tarricone et al. (2017a). The finding that standardised procedures for medical technology from the perspective of HTA institutions are not common internationally (Ciani et al., 2015; Jacobs et al., 2017) lends additional weight to the argument that South Africa would not be adopting an uncharted course in developing bespoke MedTech HTA methods but would rather be aligning itself with an emerging international consensus.

The proposed governance principles of independence, transparency and accountability align precisely with those identified by Lafortune et al. (2008) as defining the most successful HTA organisations globally. The absence of empowering legislation in South Africa represents a significant and well-documented structural gap, a problem also identified in the Colombian context by Castro (2017), who emphasised that political support and legislative architecture are prerequisites for legitimate HTA institutionalisation. Wild et al. (2017) similarly underscored the critical importance of national HTA strategy guidance and supporting legislative frameworks.

The proposed S1H2 archetype, with regulator, HTA agency, and coverage body functioning as independent entities, represents a pragmatic model for South Africa (Allen et al., 2013).

Implications for policy

The NHI presents both an opportunity and a risk for HTA institutionalisation. The opportunity lies in the political momentum and legislative vehicle that the NHI Act provides; the risk is that embedding HTA within the NHI as a transitional arrangement may compromise its independence and long-term legitimacy. A lean coordinating office that leverages existing public and private capacity is pragmatically suited to South Africa's resource-constrained environment and is consistent with the WHO survey finding that partnerships with academic institutions are useful in low-resource settings (WHO, 2015, p. 29). The determination of what will work best for South Africa (whether one or more units within the National Department of Health [NDoH] or an entirely independent body) necessarily involves trade-offs across cost, links with other health agencies, transparency, and visibility.

A critical design feature is that HTA agency recommendations should be mandatory (i.e., a positive recommendation means that the technology is adopted on essential lists and formularies), while negative recommendations should not be exclusionary (i.e., technologies surrounded by uncertainty remain available, but with conditions). Without being mandatory, recommendations risk remaining advisory only and therefore easily circumvented, undermining the entire value proposition of a national HTA system. At the same time, a clear, fair, and transparent dispute resolution process must be in place in the case of a negative recommendation, with terms of reference that are explicit and publicly available. The independence of the HTA Agency is sacrosanct, requiring a Board composed of health sector representatives nominated by relevant stakeholders, with a CEO elected by the Board and responsible to the Minister of Health.

Implications for practice

For the MedTech industry and clinical community, the proposed HTA framework carries significant practical implications. The recommendation that any stakeholder, including suppliers, may initiate a submission for HTA after market authorisation by SAHPRA provides stakeholders with a formal, equitable pathway to advocate for the appropriate assessment of innovative technologies. The tiered assessment approach (Rapid Reviews, Mini HTAs and Full HTAs) is pragmatic and avoids placing unnecessary pressure on a nascent HTA system by subjecting all technologies to the same level of scrutiny regardless of their novelty or clinical impact. The explicit inclusion of ethical, legal, social, and organisational aspects in the proposed appraisal framework represents a meaningful improvement over current practice, where medical aid schemes neither make provision for nor request such information. The EUnetHTA Core Model's nine-domain structure provides a validated international framework for this purpose (EUnetHTA, 2016). From a clinical practice perspective, clinical stakeholders, particularly clinicians, assume the role of clinical champions in the proposed model, identifying and solving implementation issues unique to their hospital institutions. A well-constituted implementation team at the healthcare institutional level, comprising people with the necessary skill sets and authority to drive multifaceted changes, is identified as essential to solve organisational problems associated with the adoption of new medical technologies.

Limitations of the study

Several limitations of this paper must be acknowledged. First, this review is written primarily from the perspective of an author with a MedTech industry background, with over forty years of experience, and therefore may also be construed as reflecting an industry-informed perspective. That said, the author is a strategy and market access consultant with HTA and health policy experience gained from working across private and public sectors. A public sector perspective is notably lacking as HTA, or practice thereof, is less developed and less transparent than in the private sector, specifically with regard to MedTech. Second, the paper does not present primary empirical data from the South African HTA context (such as a systematic mapping of existing HTA activities, their costs, timelines, and clinical or economic outcomes), which would provide a more robust evidence base for the specific structural proposals advanced. Third, the comparative analysis of European HTA archetypes, while rigorous, may not be fully generalisable to the South African context, given significant differences in health system financing, institutional capacity, and regulatory infrastructure.

Future research directions

The findings of this paper indicate several avenues for future research. First, a systematic mapping exercise of existing HTA activities across South Africa, spanning public, private, academic and professional society settings and including a skills audit, would provide the empirical baseline necessary for the institutionalisation strategy proposed. Second, primary research engaging a broad range of South African stakeholders (including patient groups, clinicians, hospital administrators, medical scheme administrators and policymakers) would enrich understanding of the political feasibility and acceptability of the proposed model. Third, research examining the impact of incorporating ELSO domains into HTA decisions on coverage outcomes for medical technologies in the South African context would generate valuable evidence for the proposed appraisal design. Fourth, the development of validated, South Africa-specific HTA methods for MedTech, drawing on the MedtecHTA Project recommendations (Tarricone et al., 2017b) and the EUnetHTA Core Model (EUnetHTA, 2016), represents a significant and time-sensitive research and policy development priority given the NHI implementation trajectory.

CONCLUSION

This paper presents a MedTech industry perspective on the establishment of an independent, transparent, and accountable HTA agency for South Africa, with a specific focus on the requirements of the medical technology sector. The central finding is that South Africa's current HTA landscape is fragmented, insufficiently transparent, and methodologically ill-suited to the unique characteristics of medical technologies. The misapplication of medicines-focused methodologies to MedTech, the absence of empowering legislation, and the dominance of cost-containment objectives over patient access considerations collectively represent critical systemic failures that an institutionalised HTA framework must address.

The proposed structural model, a lean national HTA coordinating unit supported by dedicated medicines and technology assessment units, external evaluation groups and independent multistakeholder appraisal committees, offers a pragmatic approach to HTA institutionalisation suited to South Africa's resource-constrained context (Allen et al., 2013). The proposed five-step HTA process, grounded in current practice and adapted for the specific characteristics and challenges of medical technology assessment, provides a practical and evidence-based framework for implementation.

Together, Table 1 and Figure 6, presented in the Methodology and Results sections respectively, synthesise the key process design principles and the proposed HTA process required to operationalise this framework.

The establishment of an independent HTA agency for South Africa should be underpinned by a clear legislative mandate; the sacrosanct independence of the agency from political and industry influence; mandatory and transparent recommendations; a fair and accessible dispute resolution mechanism; and equitable, tiered funding arrangements that do not place disproportionate cost burden on applicants (e.g., manufacturers). The operating costs of the agency should be funded primarily by government, with fees related to submission of dossiers being fair, centralised, and tiered according to the type of submission specific to the technology. In an era defined by South Africa's transition towards universal health coverage through the NHI, the establishment of a credible, independent, and MedTech-sensitive HTA mechanism is not merely a technical or administrative matter; it is a prerequisite for equitable patient access and the sustainable, evidence-based stewardship of health resources in a constrained environment.

REFERENCE LIST

Allen, N., Liberti, L., Walker, S. R., & Salek, S. (2017). A comparison of reimbursement recommendations by European HTA agencies: Is there opportunity for further alignment? *Frontiers in Pharmacology*, *8*, Article 384. https://doi.org/10.3389/fphar.2017.00384

Allen, N., Pichler, F., Wang, T., Patel, S., & Salek, S. (2013). Development of archetypes for non-ranking classification and comparison of European national health technology assessment systems. *Health Policy*, *113*(3), 305–312. https://doi.org/10.1016/j.healthpol.2013.09.007

Banta, D., & Jonsson, E. (2009). History of HTA: Introduction. *International Journal of Technology Assessment in Health Care*, *25*(Suppl. 1), 1–6. https://doi.org/10.1017/S0266462309090321

Battista, R. N., & Hodge, M. J. (2009). The "natural history" of health technology assessment. *International Journal of Technology Assessment in Health Care*, *25*(Suppl. 1), 281–284. https://doi.org/10.1017/S026646230909076X

Castro, H. E. (2017). Advancing HTA in Latin America: The policy process of setting up an HTA agency in Colombia. *Global Policy*, *8*(S2), 97–102. https://doi.org/10.1111/1758-5899.12333

Cheyne, S., Chakraborty, S., Lewis, S., Campbell, S., Turner, T., & Norris, S. (2023). What could health technology assessment learn from living clinical practice guidelines? *Frontiers in Pharmacology*, *14*, Article 1234414. https://doi.org/10.3389/fphar.2023.1234414

Ciani, O., Wilcher, B., Blankart, C. R., Hatz, M., Rupel, V. P., Erker, R. S., Varabyova, Y., & Taylor, R. S. (2015). Health technology assessment of medical devices: A survey of non-European Union agencies. *International Journal of Technology Assessment in Health Care*, *31*(3), 154–165. https://doi.org/10.1017/S0266462315000185

Competition Commission of South Africa. (2019). *Health Market Inquiry: Final findings and recommendations report*. https://www.compcom.co.za/wp-content/uploads/2020/01/Final-Findings-and-recommendations-report-Health-Market-Inquiry.pdf

Daubner-Bendes, R., Kovács, S., Niewada, M., Huic, M., Drummond, M., Ciani, O., Blankart, C. R., Mandrik, O., Torbica, A., Yfantopoulos, J., Petrova, G., Holownia-Voloskova, M., Taylor, R. S., Al, M., Piniazhko, O., Lorenzovici, L., Tarricone, R., Zemplényi, A., & Kaló, Z. (2021). Quo vadis HTA for medical devices in Central and Eastern Europe? Recommendations to address methodological challenges. *Frontiers in Public Health*, *8*, Article 612410. https://doi.org/10.3389/fpubh.2020.612410

Drummond, M., Griffin, A., & Tarricone, R. (2009). Economic evaluation for devices and drugs—same or different? *Value in Health*, *12*(4), 402–404. https://doi.org/10.1111/j.1524-4733.2008.00476_1.x

Drummond, M. F., Schwartz, J. S., Jönsson, B., Luce, B. R., Neumann, P. J., Siebert, U., & Sullivan, S. D. (2008). Key principles for the improved conduct of health technology assessments for resource allocation decisions. *International Journal of Technology Assessment in Health Care*, *24*(3), 244–258. https://doi.org/10.1017/S0266462308080343

European Network for Health Technology Assessment. (2016). *EUnetHTA core model version 3.0 for the full assessment of diagnostic technologies, medical and surgical interventions, pharmaceuticals and screening technologies*. http://www.htacoremodel.info/ViewHandbook.aspx

Fuchs, S., Olberg, B., Panteli, D., & Busse, R. (2016). Health technology assessment of medical devices in Europe: Processes, practices, and methods. *International Journal of Technology Assessment in Health Care*, *32*(4), 246–255. https://doi.org/10.1017/S0266462316000349

Fuchs, S., Olberg, B., Panteli, D., Perleth, M., & Busse, R. (2017). HTA of medical devices: Challenges and ideas for the future from a European perspective. *Health Policy*, *121*(3), 215–229. https://doi.org/10.1016/j.healthpol.2016.08.010

Jacobs, E., Antoine, S.-L., Prediger, B., Neugebauer, E., & Eikermann, M. (2017). Defining the relevant outcome measures in medical device assessments: An analysis of the definition process in health technology assessment. *International Journal of Technology Assessment in Health Care*, *33*(1), 84–92. https://doi.org/10.1017/S0266462317000216

Lafortune, L., Farand, L., Mondou, I., Sicotte, C., & Battista, R. (2008). Assessing the performance of health technology assessment organizations: A framework. *International Journal of Technology Assessment in Health Care*, *24*(1), 76–86. https://doi.org/10.1017/S0266462307080105

Malaysian Health Technology Assessment Section. (n.d.). *Health technology assessment manual*. Ministry of Health Malaysia. https://moh.gov.my/moh/resources/HTA_MANUAL_MAHTAS.pdf

National Department of Health. (2023). *Health technology assessment methods guide 2022–2027 to inform inclusion or exclusion of medicines for the South African national essential medicines list*. https://www.health.gov.za/wp-content/uploads/2024/04/HTA-Methods-Guide_FINAL_Sep-2023.pdf

O'Rourke, B., Oortwijn, W., & Schuller, T. (2020). The new definition of health technology assessment: A milestone in international collaboration. *International Journal of Technology Assessment in Health Care*, *36*(3), 187–190. https://doi.org/10.1017/S0266462320000215

Republic of South Africa. (2004). *National Health Act 61 of 2003*. Government Gazette No. 26595, 23 July 2004. Government Printer.

Republic of South Africa. (2024). *National Health Insurance Act 20 of 2023*. Government Gazette No. 50664, 16 May 2024. Government Printer.

Schnell-Inderst, P., Mayer, J., Lauterberg, J., Hunger, T., Arvandi, M., Conrads-Frank, A., et al. (2015). Health technology assessment of medical devices: What is different? An overview of three European projects. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, *109*(4–5), 309–318.

Siegfried, N., Wilkinson, T., & Hofman, K. (2017). Where from and where to for health technology assessment in South Africa? A legal and policy landscape analysis. In A. Padarath & P. Barron (Eds.), *South African health review 2017* (pp. 41–48). Health Systems Trust.

South African Government. (2019). *Strengthening the South African health system towards an integrated and unified health system: Presidential Health Summit compact, 25 July 2019*. https://www.presidency.gov.za/sites/default/files/2022-05/Presidential%20Health%20Compact.pdf

Tarricone, R., Torbica, A., & Drummond, M. (2017a). Challenges in the assessment of medical devices: The MedtecHTA Project. *Health Economics*, *26*(Suppl. 1), 5–12. https://doi.org/10.1002/hec.3469

Tarricone, R., Torbica, A., & Drummond, M. (2017b). Key recommendations from the MedtecHTA Project. *Health Economics*, *26*(Suppl. 1), 145–152. https://doi.org/10.1002/hec.3468

Towse, A., Devlin, N., Hawe, E., & Garrison, L. (2011). *The evolution of HTA in emerging markets health care systems: Analysis to support a policy response* (OHE Consulting Report for PhRMA). Office of Health Economics.

Wild, C., Stricka, M., & Patera, N. (2017). Guidance for the development of a national HTA-strategy. *Health Policy and Technology*, *6*(3), 339–347. https://doi.org/10.1016/j.hlpt.2017.06.006

World Bank Group. (n.d.). *South Africa* [Data table]. World Bank Open Data. Retrieved October 17, 2026, from https://data.worldbank.org/?locations=XT-ZA

World Health Organization. (n.d.). *Health technologies*. Retrieved October 17, 2026, from https://www.who.int/europe/news-room/fact-sheets/item/health-technologies

World Health Organization. (2015). *2015 global survey on health technology assessment by national authorities: Main findings*. https://iris.who.int/server/api/core/bitstreams/5fda5bbc-ea6e-407f-b6c8-5590270d5193/content

DECLARATIONS

Author contributions

Mark Brand: Conceptualisation, methodology, literature review and analysis, writing, review and editing.

Declaration of competing interest

The author is the principal of BRANDTECH Health Technology Consulting, a strategy and market-access consultancy, whose clients have historically included pharmaceutical and medtech companies, industry organisations, medical schemes, managed care and administrator companies, health practitioner groups and patient advocacy organisations. The author is an ex officio member of the South African Chapter of the Professional Society for Health Economics and Outcomes Research (ISPOR SA), a member of the QuadPara Association of South Africa (QASA), Secretary of the Southern African Spinal Cord Association (SASCA) Clinical Advisory Panel (CAP), a member of the South African Health Technology Assessment Society (SAHTAS) and an associate member of the South African Medical Technology Industry Association (SAMED). Apart from the client relationships and memberships disclosed above, the author declares no known financial conflicts of interest or personal relationships that could have influenced the work reported in this manuscript.

Declaration of ethics

The author declares that the research and writing of this manuscript adhere to ethical standards of research and publication, in accordance with scientific principles, and are free from plagiarism.

Declaration of assistive technologies in the writing process

The author declares that generative AI tools (ChatGPT and Claude) were used only in a limited and responsible manner for language refinement and structuring of sections, grammar checking and editing. The author takes full responsibility for the final content of this manuscript.



STREAM II
HTA AND PRIORITY SETTING

DEVELOPING HEALTH TECHNOLOGY ASSESSMENT–INFORMED PRESCRIBED MINIMUM BENEFIT DEFINITION GUIDELINES USING AN ADAPTED GRADE ADOLOPMENT APPROACH

Ameer Hohlfeld, Shelley McGee, Millidhashni Reddy and Trudy D Leong

How to cite this article: Hohlfeld, A., McGee, S., Reddy, M., & Leong, T. D. (2026). Developing health technology assessment–informed Prescribed Minimum Benefit definition guidelines using an adapted GRADE adolopment approach [Abstract]. *Council for Medical Schemes Research Journal*, *1*(1), 78.

ABSTRACT

Background: Prescribed Minimum Benefits (PMBs) are a regulatory mechanism to ensure access to essential health services. However, inconsistent interpretation of PMB definitions and variable use of evidence may compromise equity, efficiency and transparency in benefit design. We describe the development of evidence-informed PMB definition guidelines using an adapted GRADE adolopment approach, drawing on existing guideline evidence structured through GRADE Evidence-to-Decision (EtD) frameworks.

Methods: We applied an adapted GRADE adolopment framework, combining evidence adoption, contextual adaptation and de novo development where required. International clinical guidelines incorporating GRADE EtD tables were systematically identified. EtD domains included population eligibility, comparative clinical effectiveness, cost-effectiveness, budget impact, feasibility, equity and acceptability. Guideline quality was appraised using the AGREE II instrument, with HTA reports and economic evaluations used to supplement evidence where necessary. A multistakeholder forum, convened as a clinical advisory committee and comprising clinicians, pharmacists, HTA experts, policymakers, payers and patient representatives, reviewed the evidence and formulated PMB definition recommendations, consolidated into a structured table of recommendations through deliberation and consensus-based processes.

Results: The adapted adolopment process enabled transparent alignment between guideline-based GRADE EtD evidence, HTA findings and PMB recommendations, while permitting context-sensitive adaptation to local health system priorities and resource constraints. The resulting guidelines provide standardised criteria for defining PMB conditions, interventions and care pathways; clarify evidentiary expectations across EtD domains; and explicitly document value judgements and trade-offs. Deliberation within a multistakeholder clinical advisory committee enhanced the legitimacy, acceptability and anticipated implementability of the recommendations across regulatory and payer settings.

Conclusions: The adapted GRADE adolopment approach, leveraging existing guidelines with embedded EtD tables, supplemented by AGREE II appraisal and multistakeholder deliberation, appears to be a feasible and robust approach for developing HTA-informed PMB definition guidelines. This approach may support more transparent and consistent benefit design and be transferable to other priority-setting contexts.

KEYWORDS

Prescribed Minimum Benefits; health technology assessment; GRADE adolopment; Evidence-to-Decision framework; benefit design; South Africa



ROUTINE OUTCOMES MONITORING REPORT

BEYOND BED DAYS: CLINICAL OUTCOMES TO INFORM PSYCHIATRIST-LED STEPPED CARE AND VALUE-BASED CONTRACTING IN MENTAL HEALTHCARE

Marisa Henderson, Juanita Swanepoel and Annamarie van Jaarsveld

How to cite this article: Henderson, M., Swanepoel, J., & van Jaarsveld, A. (2026). Beyond bed days: Clinical outcomes to inform psychiatrist-led stepped care and value-based contracting in mental healthcare. *Council for Medical Schemes Research Journal*, *1*(1), 79–103.

ABSTRACT

Background: Mental healthcare systems face a persistent dilemma: rising demand, constrained specialist capacity, fragmented care pathways, variable outcomes, and escalating costs for funders and patients alike. In South Africa's private sector, this is intensified by funding structures that frequently reward service intensity more readily than clinical appropriateness. Patients sharing the same ICD-10 diagnosis may nevertheless differ substantially in psychological distress, functioning, contextual circumstances and recovery trajectories, highlighting the need for a more individualised and responsive approach to mental healthcare.

Data source and approach: This routine outcomes monitoring report draws on existing analyses of routinely collected clinical outcomes monitoring data from 9,878 adult psychiatric admissions across five Mondia Health facilities between 2022 and 2025. Psychological distress was monitored using the Depression Anxiety Stress Scale (DASS-21) at admission and discharge, with post-discharge follow-up data available for a smaller subset of patients. The article draws on data and analyses previously generated through routine outcomes monitoring, without additional data collection or statistical analyses undertaken specifically for this publication. The purpose is to reflect on observed clinical patterns and consider their implications for care delivery and service redesign.

Clinical observations and findings: Existing analyses indicated that patients with similar ICD-10 diagnoses followed different recovery trajectories depending on their psychological distress levels. Patients with moderate distress generally achieved more favourable discharge outcomes, while those with severe or extreme distress showed greater residual distress despite showing large absolute improvement scores. Follow-up data identified post-discharge vulnerability, particularly around two weeks, highlighting the importance of continued monitoring and coordinated care to sustain recovery.

Conclusions: These clinical observations suggest that while ICD-10 diagnosis identifies the condition being treated, measuring distress and functioning provides a deeper understanding of the individual patient's needs and recovery trajectory. Integrating these insights through the interdisciplinary team (IDT) enables more focused, individualised and responsive care. By making distress and recovery measurable, we can better monitor progress, guide clinical decisions, and continuously improve treatment. We therefore propose a patient-centred, psychiatrist-led stepped-care framework that aligns treatment intensity and continuity with patient needs, with the potential to improve outcomes, reduce unnecessary utilisation and manage costs more effectively. This supports a shift towards measurable, value-based mental healthcare — delivering the right care, at the right time, in the right setting.

KEYWORDS

Mental health contracting; stepped care; prognostic modifiers; transdiagnostic care; psychological distress; measurement-based care; interdisciplinary care; routine outcomes monitoring for service improvement; South Africa

INTRODUCTION

Mental healthcare in South Africa's private sector sits within a difficult and increasingly unsustainable tension: demand is rising, specialist capacity is constrained, outcomes are variable, and medical schemes face mounting pressure to contain costs while still funding appropriate care (World Health Organization, 2022). In this environment, access to care is shaped less by measured patient needs than by historical funding rules, fragmented provider arrangements, and the structure of available benefits. The result is a system in which patients frequently enter care late, at a point of acute deterioration, and schemes are asked to fund high-intensity interventions after earlier opportunities for prevention, coordinated care, and continuity have already been missed. The opposite also occurs; patients may be admitted to hospital inappropriately because admission is the only way to secure funding.

This dilemma is particularly visible in mental healthcare benefits linked to Prescribed Minimum Benefits (PMBs). South Africa's Prescribed Minimum Benefit framework, introduced in 2000, secured important protections for patients with mental health conditions, but it also formalised diagnosis-treatment pairs that are frequently interpreted in relatively fixed terms (Republic of South Africa, 1999). For example, major affective disorders were defined for hospital-based management of up to three weeks per year or outpatient psychotherapy of up to 15 contacts – or the combination thereof, while other psychiatric presentations were linked to similarly capped service packages (Republic of South Africa, 1999). While these protections secured essential coverage, they reinforced a model in which diagnosis frequently becomes the gateway to a relatively fixed package of care. Over time, this has contributed to a system that can become episodic, admission-focused, and administratively driven, rather than focused on value-based care, fully responsive to differences in severity, functioning, context, and recovery trajectory (Patel et al., 2018).

For schemes, this creates a double bind. They must fund necessary psychiatric care within a model that does not always distinguish well between patients who share a diagnosis but differ markedly in distress severity, functioning, relapse risk, social context, and likely recovery trajectory. Where care is poorly coordinated, the system becomes vulnerable not only to inefficiency, but also to fraud, waste and abuse through fragmented billing, duplication of services, repeated cycling through acute settings, and weak accountability across disconnected providers (Council for Medical Schemes, 2021).

At the centre of this problem lies an overreliance on diagnosis as the principal organiser of care intensity. ICD-10 remains essential to psychiatric practice, providing a common language for clinicians, funders, and regulators (World Health Organization, 1992). However, diagnosis alone does not adequately capture the lived burden of illness, the degree of psychological distress, the impact on functioning, or the likely course of recovery.

The prevailing mental health model focuses on specialist-led and hospital-based treatment, limiting the development of community-based care and early intervention services (Organisation for Economic Co-operation and Development, 2021). While this model ensures access to high-level care when needed, it also limits opportunities for early intervention, contributes to fragmented care pathways, and seeks early discharge as medical schemes prioritise decreased length of stay. The result is a system that is both clinically reactive and economically inefficient, with care often delivered at a stage when symptom burden and complexity have already escalated (World Health Organization, 2022).

Emerging evidence, as well as clinical observation, challenges the prevailing mental health care model. Patients with the same diagnosis frequently demonstrate substantial variability in symptom severity, psychological distress, environmental stressors, protective resources, and illness duration. These differences are not peripheral; they are structurally significant, influencing both recovery trajectories and the level of care required. Two individuals meeting criteria for the same condition may differ markedly in their response to treatment, risk of relapse, and ability to sustain recovery following discharge.

When systems fail to account for this variability, they risk applying standardised care pathways to heterogeneous populations, resulting in either under-treatment or over-servicing (Patel et al., 2018). In many physical rehabilitation settings, by contrast, care is more explicitly organised around functioning, staged intensity, and interdisciplinary planning. Patients with the same injury may receive different combinations of inpatient rehabilitation, outpatient therapy, medical review, and longer-term follow-up depending on their function and recovery goals. This distinction matters clinically and financially. When heterogeneous patients are funneled into relatively standardised pathways, some are under-treated and discharged before gains are consolidated, while others enter high-cost care only after deterioration that might have been addressed earlier through better triage and coordinated intervention. The consequence is a cycle of partial stabilisation without durable recovery, repeated utilisation, and growing expenditure for schemes and families alike (Health Quality Ontario, 2020). Mental healthcare is still too often funded and structured as though diagnosis alone should determine level of care, despite the reality that recovery is dynamic, contextual, and highly variable. This gap between diagnostic classification and lived clinical trajectory represents a critical limitation in current system design. Addressing it requires a shift from viewing diagnosis as a static label to understanding it as part of a broader, dynamic clinical formulation.

At Mondia Health, we argue for a broader clinical and funding logic in which diagnosis is complemented by prognosis at the point of assessment. This argument evolves from a single question related to health systems design: How can mental healthcare systems in South Africa be designed to encapsulate the subjective experience of mental healthcare users in its architecture, while simultaneously balancing clinical autonomy with funder realities and opportunities? In this routine outcomes monitoring report, we propose that the starting point of such a design lies in gaining a clear and coherent understanding of patient journeys. Accordingly, this report describes patterns observed through routine clinical outcomes monitoring, including differences in distress severity and improvement from admission to discharge, as well as post-discharge trajectories and readmission patterns where data were available.

Using outcomes data from 9,878 adult psychiatric admissions across five facilities nationwide, psychological distress emerged as one unified construct associated with admission-to-discharge improvement, readmission and post-discharge trajectories across varied diagnostic groups and lengths of stay. Drawing on these observed patterns, we propose a patient-centred, psychiatrist-led, interdisciplinary, stepped-care model in which comprehensive assessment of distress severity, risk, functioning, and context would inform the most appropriate entry point and treatment pathway. In such a model, the patient is at the center of an accountable interdisciplinary team, continuity is planned rather than assumed, and treatment intensity is aligned more closely with actual need. Though still untested, in proposing a stepped-care approach to mental healthcare in South Africa, we make the case for a new way of thinking: one that has the potential to enhance the current patient, practitioner and funder experience by aiming to improve outcomes, align care to need, strengthen accountability, and reduce avoidable cost.

LITERATURE REVIEW

Diagnostic classification systems remain the cornerstone of mental healthcare. Frameworks such as ICD-10 (World Health Organization, 1992) provide a shared language through which clinicians identify pathology, communicate across disciplines, and align treatment with evidence-based guidelines. The value of diagnostic clarity is fundamental, particularly in managing risk, guiding pharmacological and psychotherapeutic interventions, and ensuring consistency in clinical decision-making across settings.

However, the strength of diagnostic systems—their categorical clarity—also introduces inherent limitations. Mental health conditions are dynamic, context-dependent, and multi-dimensional in nature (World Health Organization, 2022). A diagnosis captures the presence of a condition at a point in time, but does not fully reflect its severity, variability, or likely course. This creates tension between the need for standardisation and the reality of clinical complexity as the human experience unfolds (Insel, 2022).

In practice, this limitation becomes operationally significant when diagnosis is used as the only proxy for treatment intensity. In many healthcare systems, diagnostic categories are directly linked to care pathways, benefit structures, and funding decisions (Organisation for Economic Co-operation and Development, 2021). This creates an implicit assumption that patients within the same diagnostic category require similar levels of care. Such an assumption simplifies service delivery but fails to account for heterogeneity within diagnostic groups and can produce both under-treatment and over-servicing.

Emerging evidence and clinical observation consistently demonstrate that patients with the same diagnosis may differ substantially in symptom burden, functional impairment, contextual stressors, and response to treatment. When systems rely on diagnosis alone to structure care, they risk applying standardised interventions to heterogeneous populations with implications not only for outcomes, but also for cost, continuity, and accountability, in direct contrast to value-based care.

Transdiagnostic approaches and prognostic modifiers

In response to the limitations of categorical diagnosis, transdiagnostic approaches have gained increasing prominence (Dalgleish et al., 2020). These approaches seek to identify underlying processes that cut across traditional diagnostic boundaries, recognising that many mental health conditions share common psychological and neurobiological mechanisms. Constructs such as emotional dysregulation, cognitive bias, and psychological distress have been proposed as dimensions that better capture the shared features of mental illness (Harvey et al., 2004).

Psychological distress has emerged as one robust transdiagnostic construct. The World Health Organization (2022) recognises mental health as existing along a continuum, influenced by individual, social and structural factors. Psychological distress encompasses interrelated experiences of depression, anxiety and stress, reflecting dimensions of emotional suffering that may overlap across diagnostic categories (Drapeau et al., 2012; Mirowsky & Ross, 2003). Measurement tools such as the DASS-21 operationalise this construct, enabling both cross-sectional assessment and longitudinal tracking.

However, the rise of transdiagnostic models has not been without critique. Some authors caution that an overemphasis on shared dimensions may obscure important distinctions between conditions, particularly in relation to aetiology, treatment response, and risk profiles. Diagnostic classification remains essential for identifying specific pathologies and guiding targeted interventions (World Health Organization, 2019). Without it, there is a risk of diluting clinical precision and undermining the structured frameworks that support safe practice. A more integrative perspective recognises that transdiagnostic constructs do not replace diagnosis but augment it. They provide additional layers of information that capture the intensity and lived experience of illness, complementing the categorical identification of condition. Within this view, diagnosis and transdiagnostic measures are not competing paradigms, but interdependent components of a more comprehensive clinical understanding.

Despite growing recognition of variability within diagnostic categories, the integration of prognostic factors into routine clinical formulation remains limited. This represents a critical gap between diagnostic classification and treatment planning. While diagnosis identifies the presence of a condition, it does not, in isolation, determine how that condition will evolve or what level of care will be required.

The concept of prognostic modifiers provides a framework for this shift. Prognostic modifiers include factors such as psychological distress severity, contextual stressors, protective resources, and illness duration, all of which influence the course and outcome of a condition. These factors do not replace diagnosis, but extend it, enabling a more complete understanding of patient needs.

Within this expanded framework, diagnosis and prognosis are not sequential processes, but co-defined at the point of comprehensive assessment. The moment of diagnosis becomes the moment of trajectory estimation, allowing clinicians to determine not only what condition is present, but how it is likely to unfold. From a system perspective, integrating prognostic modifiers provides a basis for stratification of care; it may support more appropriate resource allocation, help reduce unnecessary escalation and insufficient intervention, and strengthen the case for accountable, coordinated care pathways rather than isolated service events.

Measurement-based care as an enabling mechanism

The integration of prognostic modifiers into clinical practice can be operationalised through measurement-based care. This approach involves the systematic use of standardised outcome measures to assess patient status, monitor change over time, and inform clinical decision-making. By quantifying subjective experience, measurement-based care provides a structured means of incorporating transdiagnostic constructs into routine assessment (Fortney et al., 2017).

Evidence suggests that measurement-based care improves clinical outcomes by enhancing responsiveness, supporting early identification of deterioration, and facilitating timely adjustments to treatment (Fortney et al., 2017). It also promotes consistency in assessment, reducing reliance on subjective judgment alone (Lewis et al., 2019).

Beyond its clinical utility, measurement-based care serves a critical system-level function. Aggregating patient-level data enables the identification of patterns in recovery, relapse, and service utilisation (World Health Organization, 2022). This creates a bridge between individual clinical encounters and broader system design, allowing healthcare providers and funders to understand how care pathways perform over time.

In this service evaluation, routinely collected DASS-21 data describe psychological distress in relation to patient trajectories within a large clinical population. By monitoring distress across admission, discharge, and follow-up, it becomes possible to identify distinct recovery trajectories and periods of vulnerability. This provides a descriptive basis for considering how care intensity might be aligned with patient need.

Stepped care as a system response

Stepped care models have been widely proposed as a means of aligning treatment intensity with patient need (Mughal et al., 2023; Patel et al., 2018). These models are based on the principle that patients should receive the least intensive effective intervention, with escalation to higher levels of care when clinically indicated (Bower & Gilbody, 2005; Mareya et al., 2024). In theory, stepped care offers a framework for improving access, optimising resource utilisation, and enhancing continuity.

However, the implementation of stepped care has varied across settings, reflecting differences in its definition and application (Mughal et al., 2023).

Without reliable indicators for differentiating patient need, stepped care risks becoming another form of standardisation, with patients progressing through predefined steps rather than being matched to appropriate levels of care. In such cases, the model fails to fully realise its potential.

The integration of prognostic modifiers provides a means of strengthening stepped care by introducing a clinically grounded basis for differentiation. By incorporating measures of distress, contextual factors, and illness trajectory into initial assessment, patients may be more appropriately stratified and directed to appropriate levels of care from the outset. This could help reduce unnecessary escalation, prevent premature discharge, and support continuity across the treatment pathway. Within this framework, stepped care is not merely a model of service delivery, but a translation of clinical insight into system design that supports patient-centred care. It operationalises the relationship between diagnosis, prognosis, and treatment intensity, creating a structure that is responsive to patient trajectories rather than constrained by categorical definitions alone.

Importantly, this has direct implications for healthcare funding. In systems where cost is driven by high-intensity interventions, particularly inpatient care, the ability to align treatment with need has significant economic value (World Health Organization, 2022). By reducing reliance on costly services, improving continuity, and preventing relapse, stepped care offers a potential pathway toward more efficient and sustainable mental healthcare delivery.

CLINICAL OBSERVATIONS AND FINDINGS

Outcomes monitoring rationale

This outcomes monitoring report is based on routinely collected and retrospectively reviewed clinical outcomes data generated as part of clinical care and quality monitoring. The work was undertaken for service improvement and was not designed or conducted as research. The primary aim of the outcomes-monitoring process was to examine patterns of recovery across psychiatric admissions and to explore the relationship between psychological distress and patient trajectories within real-world clinical settings, to inform service improvement in the Mondia Health network.

The retrospective nature of outcomes monitoring enables system-level pattern identification, rather than controlled intervention effects; and the examination of longitudinal trends across a substantial period (2022–2025). This provides governance level insight into consistency and variability in outcomes across the service provider network. The analyses presented in this report were undertaken as part of Mondia Health's existing outcomes-monitoring and service-improvement processes, with no additional data collection or statistical analyses conducted specifically for this publication.

Data were extracted for the period January 2022 to mid-2025, providing a multi-year dataset that captures variation across time and service contexts. Data were obtained from the Mondia Health clinical registry which includes standardised clinical assessments conducted as part of routine care, allowing for consistency in data collection across sites.

Mondia Health facilities provide inpatient and outpatient adult mental healthcare services across a range of psychiatric conditions, including mood disorders, anxiety disorders, substance-related disorders, and other mental health conditions.

Patient population and record inclusion criteria

The number of patient records utilised for outcomes monitoring during the service evaluation period consisted of 9,878 unique adult psychiatric admissions across five Mondia Health facilities in South Africa. Case records were eligible for analysis when:

- A patient aged 18 years or older provided written informed consent for data outcomes use as per the service provider’s established processes.

- Records showed completed DASS-21 assessments on admission and discharge.
- ICD-10 codes reflected any of the following diagnoses: major depressive disorder, bipolar mood disorder and/or substance-use disorder.

The reported population of 9,878 represents the number of unique patients rather than admission episodes. Analyses of changes in distress were based on the available assessment records relevant to each comparison.

A subset of patients (n=270) who completed follow-up assessments at two weeks and six weeks post-discharge was included in the longitudinal analysis of post-discharge trajectories.

The dataset reflected routine clinical admissions within the specified diagnostic groups. Cases with identifiable psychosis and/or other reality-disorientating conditions such as dementia were excluded.

DATA POINTS AND METRICS

Psychological distress (DASS-21)

Psychological distress was measured using the Depression Anxiety Stress Scale (DASS-21), a validated self-report instrument widely used in clinical and research settings. The Depression Anxiety Stress Scales (DASS) were originally developed by Lovibond and Lovibond (1995) as a comprehensive measure of negative emotional states, with the shortened 21-item version (DASS-21) gaining popularity due to its efficiency and strong psychometric properties. Subsequent research has consistently supported its reliability and validity across diverse populations. For instance, Lee et al. (2019) conducted a systematic review using updated COSMIN methodology and found evidence supporting several measurement properties of the DASS-21, while identifying areas requiring further validation.

Regional validations further reinforce its utility; Dreyer et al. (2019) found the instrument to be reliable and valid in a South African working adult population, providing further evidence of its validity in a South African non-clinical population. Collectively, these findings establish DASS-21 as a widely used, validated self-report tool in both research and clinical practice. The DASS-21 consists of three subscales measuring depression, anxiety, and stress, each contributing to an overall measure of distress and capturing distress levels as a continuous, measurable dimension, reflecting symptom intensity (Lovibond & Lovibond, 1995).

Scores were calculated according to standard guidelines and categorised into severity bands: normal, mild, moderate, severe, and extremely severe. These categories were used to stratify patients and examine differences in recovery trajectories. The routine clinical use of the DASS-21 offers several advantages for outcomes monitoring. First, its transdiagnostic nature allows for comparison across diagnostic categories. Second, it provides a quantifiable measure of subjective experience, which is central to focus on prognostic modifiers. Third, its routine use within the clinical setting ensured high data availability and consistency.

Other data points and metrics

Diagnostic information was recorded according to ICD-10 criteria as part of routine clinical assessment. While diagnosis was not the primary variable of interest, it provided essential context for interpreting patterns across patient groups. Length of stay (LOS) was calculated as the number of days between admission and discharge for each inpatient episode. This variable was used to examine variation in treatment duration across diagnostic categories and patient groups.

Readmission was defined as any subsequent inpatient admission following discharge within a defined 30-day period post-discharge. The reported readmission proportion was calculated as the percentage of unique

patients who experienced at least one readmission during this period. It is important to note that readmission rates mentioned, refer to traceable readmissions of patients within the single provider network and does not account for admissions in other facilities of which we may not be aware. For patients with available data, DASS-21 scores at two weeks and six weeks post-discharge were used to examine post-discharge trajectories and identify periods of vulnerability. These post-discharge data sets were collected telephonically by the patient's attending therapist or therapy team member during a follow-up check-in scheduled 2- and 6-weeks post discharge.

It is important to note that the authors acknowledge that while this outcomes monitoring report focused on psychological distress, other transdiagnostic constructs such as duration of illness, protective resources and environmental factors may have acted as confounding variables in the trajectories discussed.

Analysis

Data were analysed by an independent statistician not affiliated to the service provider network. Data analysis was conducted in several stages, each aligned with outcomes monitoring objectives. Stage one required the composition of an overall picture of patient profiles from admission to discharge, including diagnostic categories, LOS data, and distress levels on admission and discharge to track improvement/regression trajectories. Stage two involved clearing up the data to a) reduce any duplications, b) exclude incomplete and confounding data sets, and c) control for continuous care across recovery and wellness wards, leading to longer admission periods. Stage three focused on the identification of factors associated with patient outcomes and recovery trajectories across the group.

Correlation analysis, with relationships between subscales not assumed but empirically tested, was used to examine relationships between DASS-21 subscales. Principal component analysis (PCA) was conducted to assess whether a single underlying factor could account for variance across subscales. This step informed the interpretation of psychological distress as a common dimension across the observed symptom measures. Patients were grouped according to diagnosis and distress severity at admission using ICD-10 and standard DASS-21 thresholds. This stratification enabled comparison of recovery trajectories across diagnostic groups and distress severity levels. Changes in distress severity between admission and discharge were examined within each severity group. For patients with follow-up data, trajectories were extended to include two-week and six-week time points. Length of stay and readmission rates were analysed descriptively, with comparisons made across diagnostic categories and severity groups. These analyses not only added temporal depth but also provided insight into how clinical patterns interact with system-level variables.

The analytical approach was primarily descriptive and exploratory. The aim was not to establish causal relationships, but to identify patterns that could inform clinical and system-level understanding. Given the observational nature of the data, findings should be interpreted as indicative rather than definitive. Findings from our analysis do not claim that distress causes specific outcomes, but rather that it is associated with distinct patterns of recovery and utilisation. This distinction is important in maintaining methodological rigor. While the patterns observed are robust and consistent, they should be understood as informing service improvement and system design, rather than as proof of causal mechanisms.

Ethical considerations

This article reports a service evaluation based on outcomes data collected routinely as part of clinical care, outcomes monitoring and service improvement at Mondia Health. The work was not designed or conducted as research. No additional interventions, data collection or statistical analyses were undertaken specifically for this publication. All assessment records form part of routine clinical care and are used to support clinical assessment, goal contracting and discharge reporting for individual patients.

Patients consented to the collection, recording and appropriate use of their routinely collected clinical information for outcomes monitoring and service improvement through established admission procedures.

Consent is generally obtained by a qualified nursing practitioner during patient admission. Where a patient is temporarily unable to provide informed consent, consent is deferred until the patient can understand and agree to the relevant conditions. Mondia Health admits voluntary adult patients only, and no proxy consent is obtained for outcomes monitoring. Where the required consent is not obtained, the relevant data are excluded from the evaluation. Patients who decline consent experience no penalty or prejudice in their clinical care. Routine post-discharge assessments at two and six weeks formed part of the established clinical follow-up process and were conducted in accordance with the applicable patient consent arrangements.

The evaluation used de-identified clinical data, and all findings are reported in aggregate form, without information that could identify individual patients. Routine clinical records are securely stored and managed within the hospital network in accordance with organisational data-protection procedures and the Protection of Personal Information Act 4 of 2013 (POPIA). Publication of these de-identified, aggregate findings does not disclose identifiable patient information.

Mondia Health's routine outcomes reports inform data-driven clinical decision-making and ongoing service improvement. These activities include the refinement of clinical systems and procedures, staff training, adjustments to therapeutic programmes, and communication of aggregated service outcomes to relevant stakeholders, including treating physicians, psychologists, healthcare funders and health policymakers. These activities form part of Mondia Health's established clinical governance, quality monitoring and service-improvement processes.

RESULTS

Population characteristics and data completeness

The service-monitoring population comprised 9,878 unique adult psychiatric patients across five facilities between January 2022 and mid-2025. The reported analysis focused on patients with primary diagnoses of major depressive disorder, bipolar mood disorder and substance-use disorder, as classified according to ICD-10 criteria.

DASS-21 assessments were routinely administered at admission and discharge, and available assessment records were used for the reported analyses. A subset of 270 patients completed follow-up assessments at two weeks and six weeks post-discharge, allowing for limited longitudinal analysis of post-discharge recovery patterns. While follow-up data were not available for all patients, the available subset was sufficient to identify consistent trends in post-discharge trajectories.

Structure of psychological distress

The results show substantial overlap in admission distress severity across diagnostic groups, with a strong general distress factor dominating the finer distinctions. This finding resulted from stage three of the data analysis and met the objective of identifying possible unifying constructs linked to the subjective experience of patients' mental health conditions.

Analysis of DASS-21 subscales demonstrated strong positive correlations between depression, anxiety, and stress scores, with correlation coefficients ranging from 0.69 to 0.78. Principal component analysis further indicated that a single dominant component accounted for approximately 78% of the variance across subscales. This finding is consistent with continental evidence demonstrating high co-morbidity and shared emotional underpinnings among psychiatric populations.

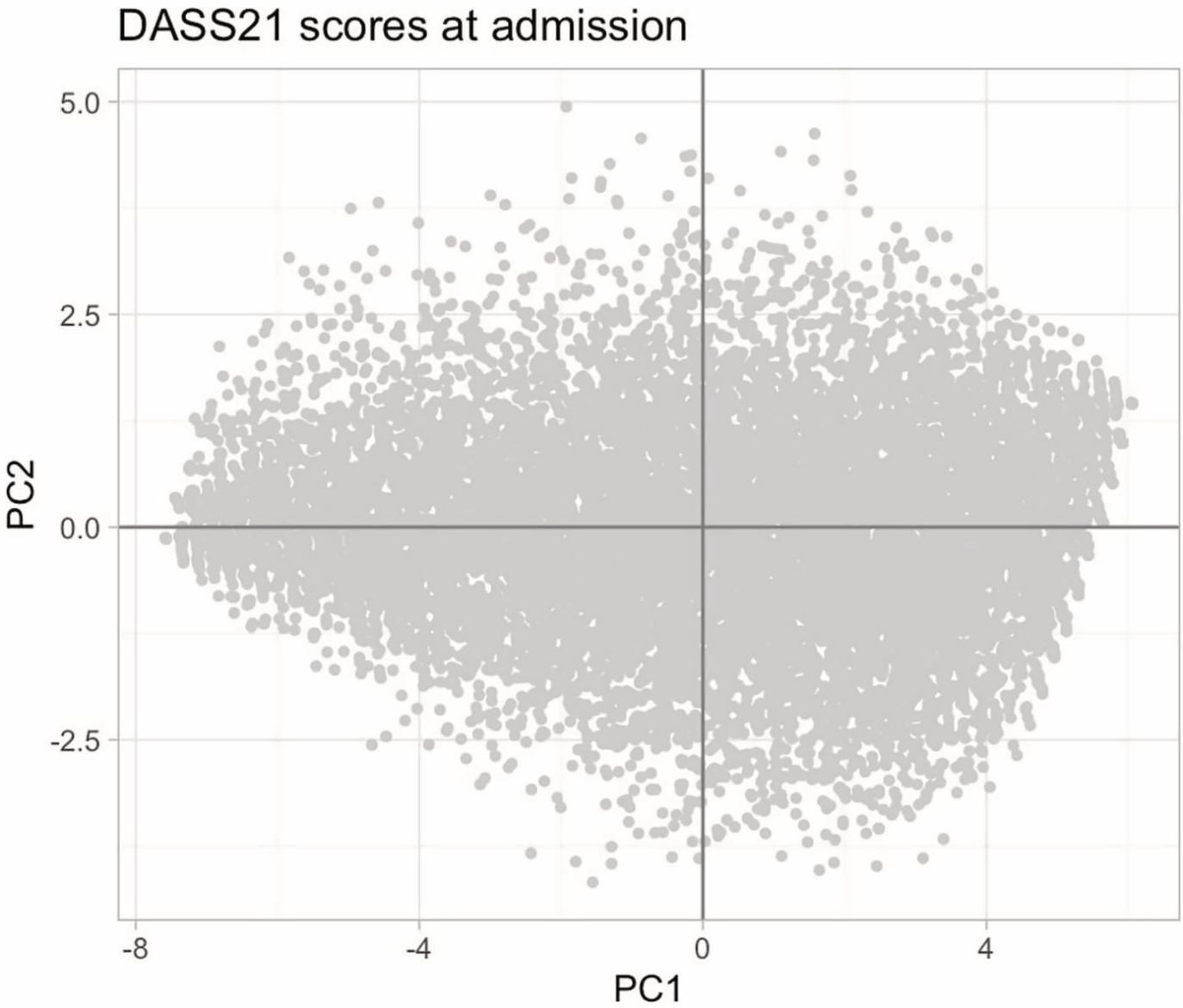

Figure 1: DASS-21 scores on admission, combined across subgroups (n = 9,878; Mondia Health, 2025)

Figure 1 illustrates the general distress factor derived from routinely collected DASS-21 admission scores for the evaluated patient population between 2022 and 2025.

The plot in figure 2 shows that the first principal component (PC1), which corresponds to the sum of the items, i.e. reflects the total DASS21 score, explains most of the variance in patient experience. Secondly, the plot shows that depression, anxiety and stress scores are highly correlated (r = 0.69, 0.73 and 0.78). Consequently, patients tend to score high or low across all dimensions of the DASS-21. Distress influences the subjective experience of a patient's wellbeing and cannot only be ascribed to a single area of life being affected. In cases like these, the DASS-21 indicates the measure of a general distress factor.

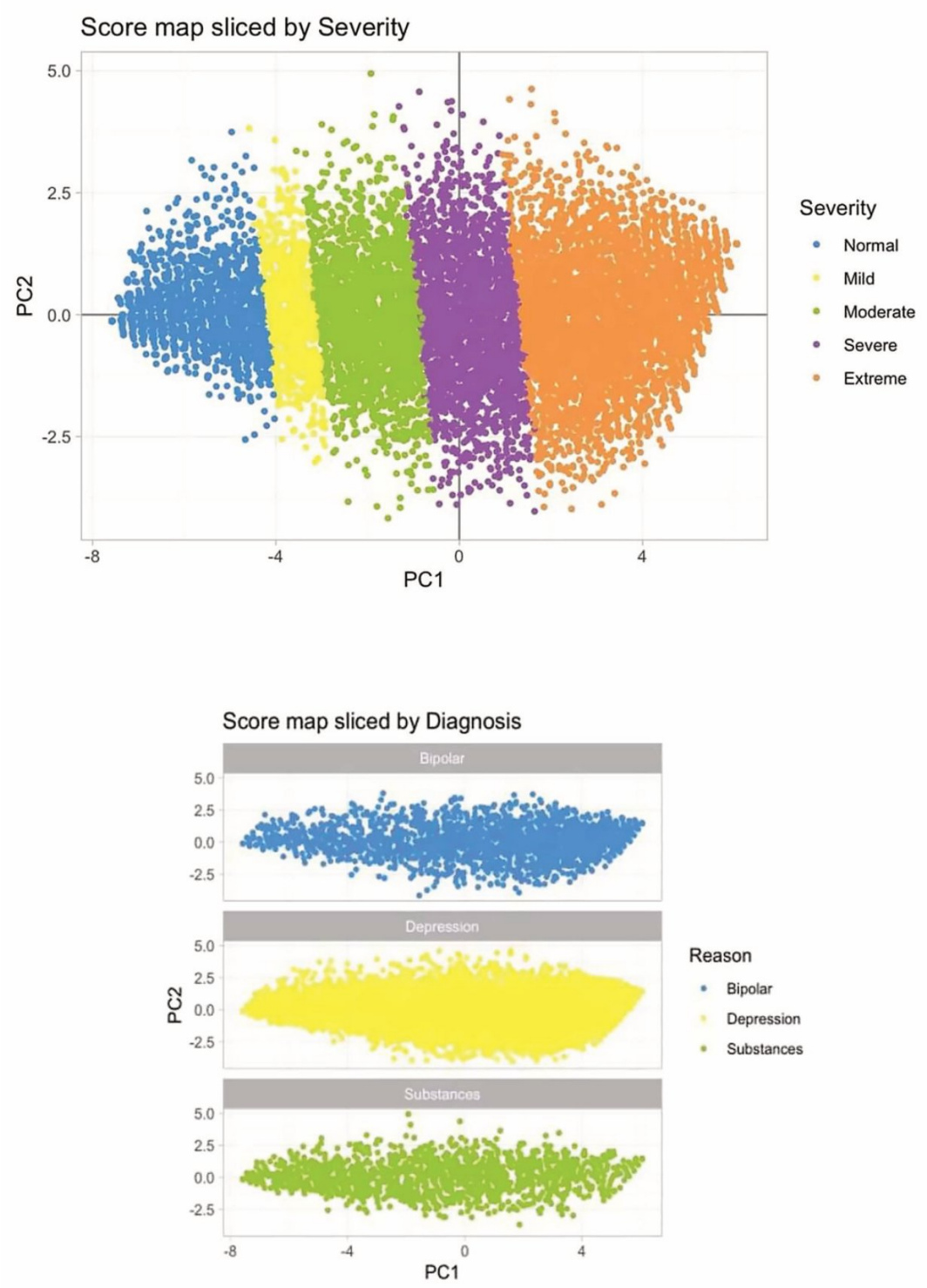

Figure 2: DASS-21 scores on admission, by subgroup (n = 9,878; Mondia Health, 2025)

Principal-component analysis of the sliced scores, confirmed that a single general-distress factor accounts for most variance across diagnostic subgroups, rather than discrete symptom clusters. This means that when the general distress plot is split into diagnosis-specific plots, substantial overlap remains in the distribution of depression, anxiety and stress scores across diagnostic groups. The diagnosis does not account for the variance in the patient experience of depression, anxiety and stress. Therefore, generalised distress, as measured by the sum of the DASS-21 scores, best explains any variance in severity on admission and subsequently describes the core, subjective, emotional experience of our patients best. These findings support the interpretation of psychological distress as a unified construct within this population. For the purposes of subsequent analyses, distress severity was therefore treated as a composite measure, with patients categor-

ised into standard severity bands (normal, mild, moderate, severe, and extremely severe) based on DASS-21 thresholds.

Distress severity at admission and discharge

Changes in distress severity between admission and discharge were examined across the full sample during stage one and two of the data analysis and met the objective of tracking entry and recovery patterns across the time of admission to discharge. Overall, patients demonstrated reductions in distress levels during admission, with shifts observed from higher severity categories toward lower categories at discharge. However, when stratified by admission severity, distinct patterns emerged. A greater proportion of patients presenting with moderate distress reached the normal range at discharge compared with those presenting with severe or extremely severe distress. In contrast, patients presenting with severe or extremely severe distress demonstrated reductions in symptom severity but were less likely to reach normative levels at discharge.

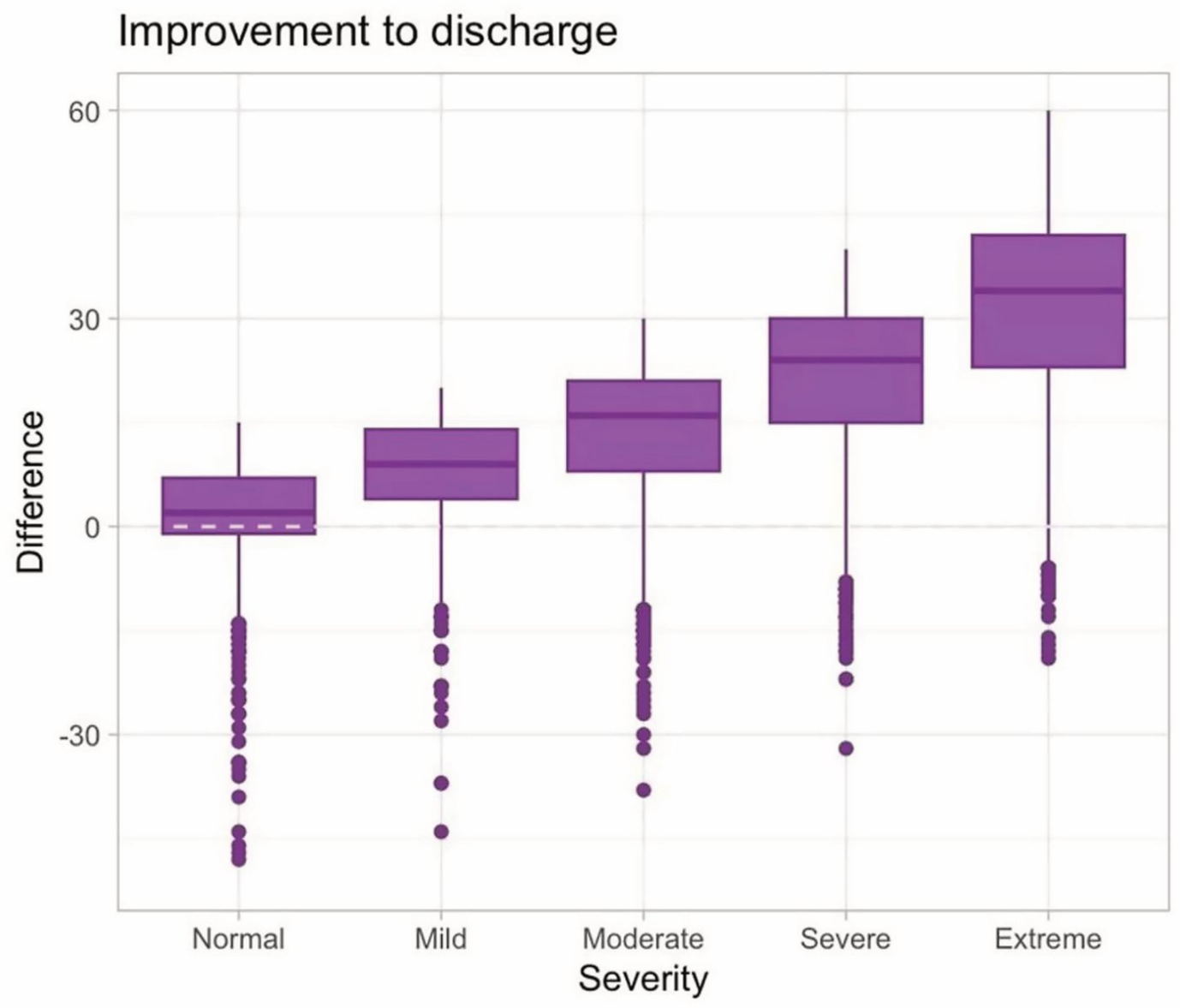

Figure 3: Improvement in distress levels from admission to discharge (n = 9,878; Mondia Health, 2025)

Figure 3 illustrates the distribution of changes in DASS-21 scores between admission and discharge, grouped according to admission severity. This figure demonstrates that while improvement occurred across all groups, the extent of recovery varied substantially depending on initial severity.

Recovery trajectories by severity group

To further examine these patterns, patients were grouped according to their distress severity at admission, and changes in severity were tracked across admission and discharge.

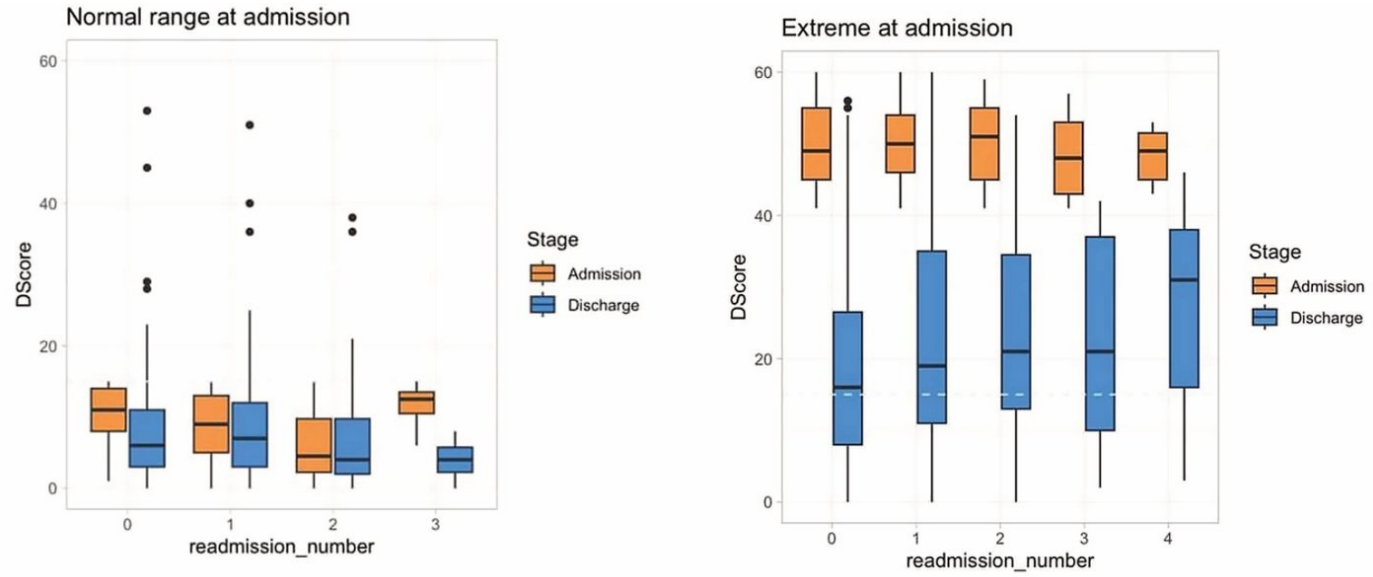

Figure 4: Admission and discharge DASS-21 scores across successive admissions among patients initially presenting with normal or extreme distress (Mondia Health, 2025).

Patients in the moderate severity group demonstrated favourable discharge outcomes, with 71% reaching the normal range. Patients admitted with severe and extreme distress showed larger absolute improvements, although smaller proportions reached the normal range at discharge (64% and 54%, respectively). Figure 4 illustrates admission and discharge DASS-21 scores across successive admissions for patients initially presenting with normal or extreme distress. The patterns demonstrate improvement during individual admissions, while distress levels at subsequent readmissions frequently resemble those recorded during earlier admissions.

Length of stay across diagnostic groups

Length of stay (LOS) was analysed across diagnostic categories as part of routine outcomes monitoring to describe treatment duration and patterns of service utilisation within each diagnostic group. Length of stay was concentrated around 10 days for patients with major depressive disorder, varied among patients with bipolar mood disorder, and was generally longer for patients with substance-use disorders, with the distribution peaking at approximately 20 days.

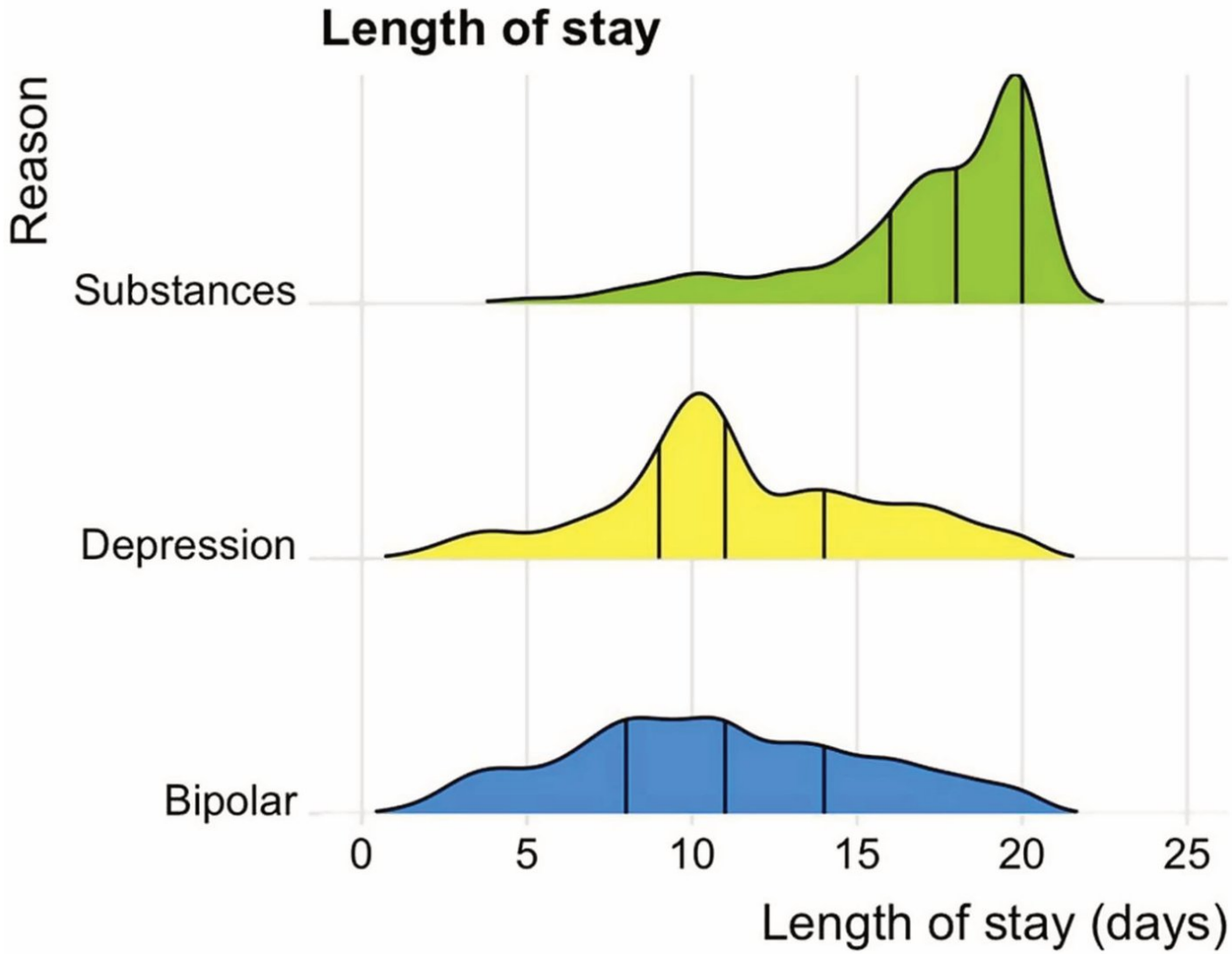

Figure 5: Length of stay for substance-use disorder, major depression and bipolar mood disorder (n = 9,878; Mondia Health, 2025)

The distribution of LOS by diagnostic category is presented in Figure 5. This figure illustrates variability in treatment duration across conditions, indicating that LOS is not uniform across patient groups.

Readmission patterns

Readmission patterns were examined as part of routine outcomes monitoring to identify periods of vulnerability following discharge. Of the 9,878 unique patients, 18% were readmitted at least once within the Mondia Health network during the evaluation period. Among these patients, distress levels at readmission were frequently comparable to those observed at initial admission.

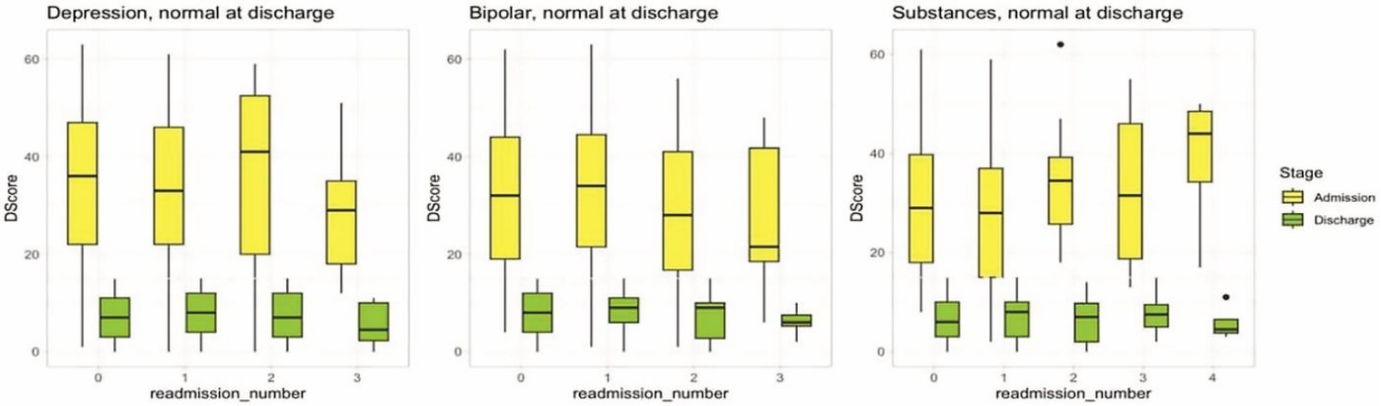

Figure 6: Admission and discharge DASS-21 scores across successive admissions among patients discharged within the normal distress range, by diagnostic group (Mondia Health, 2025).

Figure 6 illustrates admission and discharge DASS-21 scores across successive admissions among patients previously discharged within the normal distress range. The patterns show that patients may return for subsequent admissions with distress levels comparable to those recorded during earlier admissions, despite achieving normal-range scores at discharge.

Post-discharge distress trajectories

For patients with available follow-up data (n = 270), distress levels were examined at two weeks and six weeks post-discharge. Among patients discharged in the normal, mild and moderate bands, distress scores increased at two weeks after discharge and continued to rise between two and six weeks, whereas scores for patients discharged with severe or extreme distress did not increase at two weeks. The increase at two weeks indicates a period of vulnerability following discharge, during which patients may be at increased risk of relapse or deterioration.

The findings highlight the continued vulnerability of patients even when discharged within the normal range. The increase in distress at two weeks, continuing to six weeks in some severity groups, highlights the importance of responsive, well-coordinated post-discharge care.

Figure 7 demonstrates the differences in post-discharge patterns between normal and extreme severity groups.

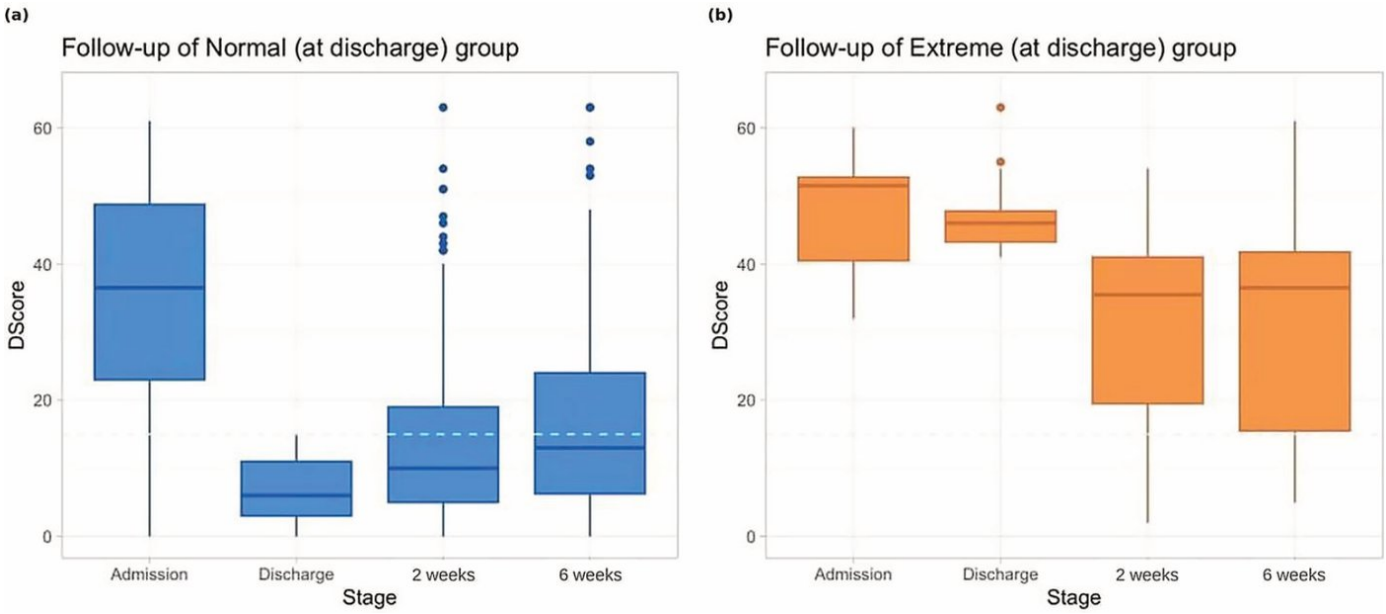

Figure 7: Differences in post-discharge patterns between (a) normal and (b) extreme severity groups (Mondia Health, 2025)

The post-discharge distress patterns of patients discharged within the extreme distress range suggest a need for extended care and coordinated aftercare to address their continued vulnerability, particularly where distress remains elevated at discharge. This aligns with the WHO's emphasis on continued and community-based mental healthcare and support (World Health Organization, 2022); a need also highlighted by South African findings of limited aftercare infrastructure and inconsistent community linkages in mental healthcare (Firfirey-Brijlal et al., 2025; Gamieldien et al., 2025).

The results demonstrate three consistent patterns. First, psychological distress functions as a unified construct that can be used to stratify patients across severity levels. Second, recovery trajectories differ substantially depending on distress severity at admission, with patients in higher severity categories showing incomplete recovery at discharge. Third, post-discharge vulnerability and readmission patterns indicate that improvements achieved during admission are not consistently sustained.

Together, these findings provide a descriptive account of variability in patient trajectories across the treatment pathway.

DISCUSSION

The findings of the outcomes monitoring report illuminate a central tension within contemporary mental health systems: the persistent reliance on diagnostic classification as the primary determinant of care intensity, despite observations suggesting that diagnosis alone does not adequately account for differences in patient recovery trajectories. While diagnostic systems such as ICD-10 remain indispensable for identifying pathology, guiding treatment, and maintaining clinical coherence, the data presented here demonstrate that they do not fully account for the variability observed in recovery patterns.

This tension is not merely theoretical. It manifests concretely in how care is delivered, funded, and experienced. When systems treat diagnosis as a proxy for need, they risk applying uniform treatment pathways to inherently heterogeneous patient populations. The findings provide empirical support for this concern. Patients with similar diagnostic classifications demonstrated markedly different recovery trajectories when stratified by distress severity, suggesting that diagnosis alone is insufficient to guide decisions regarding treatment intensity, duration, and follow-up care.

The implication is not that diagnosis should be diminished or replaced. On the contrary, the findings reinforce its centrality. However, they also indicate that diagnosis must be understood as incomplete when considered in isolation. The concept of prognostic modifiers provides a framework for addressing this limitation. By integrating factors such as distress severity, contextual stressors, protective resources, and illness duration into clinical formulation, it becomes possible to move beyond static categorisation toward a more dynamic understanding of patient need.

Among these modifiers, psychological distress emerges as particularly significant. The analysis confirmed that distress functions as a unified construct, with strong correlations across depression, anxiety, and stress domains. This supports the broader transdiagnostic literature, which has increasingly recognised that emotional disorders share common underlying processes. However, the value of distress lies not only in its conceptual coherence, but in its practical utility as a predictor of trajectory.

When patients were grouped according to distress severity at admission, distinct patterns of recovery became visible. Patients presenting with moderate levels of distress frequently reached the normal distress range by discharge. In contrast, those presenting with severe or extremely severe distress demonstrated larger absolute improvements during admission but were less likely to reach the normal range at discharge. Post-discharge patterns varied according to discharge severity. This pattern suggests that distress severity functions as a prognostic gradient, differentiating not only the starting point of illness, but its likely course.

This finding deepens the argument introduced in the literature overview regarding the limitations of diagnosis- only models. It demonstrates, within a real-world dataset, that two patients with the same diagnosis may require fundamentally different levels of care based on their position along this gradient. It also highlights the importance of timing. The observed increase in distress approximately two weeks post-discharge among patients discharged within the normal, mild and moderate ranges, highlights a potential period of vulnerability. These findings suggest that post-discharge monitoring and continuity of care remain important, even among patients with favourable discharge outcomes. Patients with higher levels of distress at discharge appear less able to sustain gains during this period, suggesting that the interaction between severity and continuity is critical.

This interaction points to a second, equally important limitation in current systems: the fragmentation of care across the treatment pathway. In the prevailing model, inpatient admission functions as a discrete episode, with limited integration into ongoing care. Stabilisation is achieved within a controlled environment, but the transition back into the community is often insufficiently supported. This discontinuity disrupts the trajectory of recovery, particularly for patients whose distress levels remain elevated at discharge.

The readmission data reinforce this interpretation. The finding that 18% of patients were readmitted, often with distress levels returning to baseline, suggests that initial improvements were not consolidated. Rather than representing new episodes of illness, these readmissions may reflect incomplete recovery within the original episode.

From a clinical perspective, this raises questions about whether the intensity and duration of care were appropriately aligned with patient need. From a system perspective, it highlights the inefficiency of repeatedly stabilising patients without addressing the underlying trajectory.

At this point, the relationship between clinical variability and system design becomes unavoidable. The data do not simply describe patient behaviour; they reveal the consequences of structural assumptions embedded within the system. When care is organised around fixed durations, standardised pathways, and discrete episodes, it cannot adequately respond to the dynamic nature of mental illness. The result is a mismatch between patient trajectories and service delivery, which manifests as both suboptimal outcomes and increased cost.

This mismatch is particularly evident in the use of inpatient care. The analysis of length of stay indicates that different patient groups require different durations of treatment. Mood disorder admissions averaged approximately ten days, while substance-related admissions extended to approximately twenty days. These findings stand in contrast to standardised benefit structures, which often allocate fixed durations regardless of clinical variation. Such structures implicitly assume that patients within a category are similar enough to warrant identical treatment durations, an assumption that the data clearly challenge.

From a clinical standpoint, this rigidity can lead to both under-treatment and over-servicing. Patients requiring longer or more structured care may be discharged prematurely, increasing the risk of relapse. Conversely, patients who stabilise more quickly may remain in inpatient care longer than necessary, exposing them to unnecessary intervention and contributing to inefficiency. From a financial perspective, this rigidity amplifies cost by tying expenditure to predefined structures rather than actual need.

The substantial resource requirements associated with inpatient care further sharpen this argument. Bed-days represent the most significant component of mental health expenditure, and their utilisation is strongly influenced by how the system is structured. When inpatient care is positioned as the default response, it becomes both a clinical and financial anchor. Any effort to improve efficiency must therefore address not only the cost of beds, but the pathways that lead patients into them. It is at this intersection—between clinical variability, system design, and cost—that the concept of stepped care gains its relevance. Our findings do not merely support stepped care as a theoretical model; they provide the empirical basis for its necessity. By demonstrating that patient trajectories differ according to measurable factors, the data create the conditions for differentiation. Stepped care then becomes the mechanism through which this differentiation is operationalised.

However, it is important to distinguish between stepped care as a general concept and its specific application in this context. In many implementations, stepped care is framed primarily as a means of improving access or reducing waiting times. While these are important objectives, they do not fully capture the implications of the present findings. Here, stepped care is positioned as a response to the inadequacy of diagnosis-only models and the inefficiency of fragmented systems. Its purpose is not simply to distribute care more widely, but to align it more precisely with patient trajectories.

This alignment requires two critical shifts. The first is conceptual: moving from diagnosis as a static label to diagnosis as a dynamic formulation that includes prognostic modifiers. The second is structural: reorganising care pathways to reflect this formulation. Together, these shifts enable a system in which the intensity, duration, and continuity of care are determined by patient's need rather than predefined categories.

The role of psychiatry within this model also warrants careful consideration. The proposed approach does not diminish the importance of psychiatrists; rather, it repositions their role within a broader system. By maintaining psychiatric leadership while expanding access through lower-intensity tiers and inter-disciplinary collaboration, the model seeks to optimise the use of specialist expertise. Psychiatrists remain central to diagnostic formulation, risk management, and complex care, but are no longer the sole entry point into the system.

For clinicians, the proposed model offers a framework for integrating diagnosis with measurable distress, functioning and other prognostic indicators to better understand patient trajectories and guide individualised treatment through the interdisciplinary team (IDT). For medical schemes, aligning care intensity and continuity with patient needs presents an opportunity to improve resource allocation, reduce potentially avoidable inpatient utilisation and readmissions, and manage costs more effectively. These anticipated clinical and economic benefits require further evaluation.

These clinical observations highlight a broader challenge in mental healthcare: aligning individual patient needs with available services and funding structures. Integrating prognostic indicators into diagnostic formulation, supported by coordinated interdisciplinary care and ongoing outcome measurement, may provide a practical foundation for more responsive, effective and financially sustainable mental healthcare. The following section presents a proposed stepped-care framework to translate these principles into practice.

PROPOSED MODEL: A PSYCHIATRIST-LED STEPPED CARE MODEL

The stepped care model proposed in this report is not introduced as an abstract concept, but as a direct response to the patterns observed in the data. The differentiation in recovery trajectories, the variability in length of stay, and the persistence of readmissions collectively point to a system that is insufficiently responsive to patient need. The model therefore seeks to translate these insights into a structured approach to care delivery. Figure 8 presents the stepped care model and highlights distress-based stratification, dynamic stepped allocation and inter-disciplinary care, care intensity and setting, and trajectory-based funding as its core components.



Figure 8: Components of the proposed psychiatrist-led stepped-care model (authors' conceptual illustration, informed by Mondia Health, 2025).

The model is built on the principle of intensity alignment. Care is positioned at the level required by the patient's current state and adjusted dynamically as that state changes. This requires both accurate initial stratification and the capacity for ongoing reassessment. It also requires a system architecture that supports movement between levels of care without fragmentation.

The model is psychiatrist-led, ensuring that specialist expertise remains central to clinical governance. However, it expands the points of access and redistributes care across a continuum, rather than concentrating it within inpatient settings. This redistribution is not a reduction in care, but a reconfiguration of where and how

care is delivered. Table 1 presents a comparison between the traditional fee-for-service (FFS) model currently widely adopted, and the proposed stepped care model.

Table 1: Comparison between the traditional fee-for-service model and the proposed stepped-care model

Component	Fee-for-service model	Stepped-care model
Entry point	Specialist or admission	GP plus structured triage for early intervention; psychiatrist-led escalation and risk management
Bed utilisation	High	Targeted and aligned with patient need
Length of stay	Fixed	Needs-based, data-driven packaging
Readmissions	Higher	Reduced through improved continuity of care after discharge and flow between tiers
Care coordination	Fragmented	Integrated, inter- and multidisciplinary
Primary cost driver	Inpatient beds	Distributed across tiers and aligned with intensity of care

Tier A: Early intervention and system entry

Tier A exists to address the most fundamental bottleneck in the current system: delayed access to care. In the existing model, patients often enter treatment only once symptoms have escalated to a level requiring specialist intervention. This delay not only increases clinical complexity but also drives cost, as higher-intensity interventions become necessary.

Tier A provides a structured entry point into care at a lower level of intensity, supported by primary care practitioners and guided by psychiatric oversight. Its purpose is not merely to treat mild conditions, but to intercept trajectories before they escalate. The theoretical basis for this tier is grounded in both early intervention literature and the findings that patients with lower levels of distress demonstrate more favourable recovery trajectories. By expanding access at this level, Tier A reduces pressure on higher tiers, prevents unnecessary escalation, and shifts the system from reactive to proactive care.

Tier B: Diagnostic-prognostic integration

Tier B represents a critical innovation in the model. It responds directly to the finding that diagnosis alone does not sufficiently guide care planning. This tier provides a structured environment in which diagnostic classification is integrated with prognostic modifiers to produce a comprehensive clinical formulation. The purpose of Tier B is not simply to confirm diagnosis, but to define trajectory. By incorporating distress severity, contextual factors, and illness history, clinicians can determine the most appropriate pathway for each patient. This reduces the risk of both under-treatment and over-servicing. From a system perspective, Tier B functions as a triage mechanism. It ensures that patients are directed to the level of care most suited to their needs, thereby improving both outcomes and efficiency.

Tier C: Intensive outpatient stabilisation

Tier C is designed for patients with significant distress and functional impairment who do not require full inpatient admission. The report's findings indicate that many patients experience incomplete recovery during standard inpatient stays, suggesting that the issue is not always the absence of care, but the misalignment of its form and duration. Tier C provides an alternative by delivering high-intensity interventions in an outpatient or partial hospitalisation setting. This allows for greater flexibility in duration and maintains patient connection to their community environment. It also reduces reliance on inpatient beds, which are the primary cost driver in the system.

Tier D: Acute inpatient care

Tier D retains the essential role of inpatient care for patients with acute risk or severe complexity. However, within the stepped care model, inpatient care is repositioned as one component of a broader continuum, rather than the central organising principle of the system. Its function is clearly defined: containment, stabilisation, and preparation for step-down into lower tiers. This integration reduces the likelihood of discharge into unsupported environments and mitigates the risk of readmission.

The role of medical schemes in stepped care

Medical schemes face increasing pressure to contain healthcare expenditure, with cost management often focused on limiting inpatient bed days and services provided. However, the fragmented nature of mental healthcare creates additional administrative costs through the separate authorisation, monitoring and auditing of hospitals and individual service providers, often without a consolidated view of patient outcomes. While utilisation controls are necessary for financial sustainability, focusing primarily on individual services and short-term expenditure may inadvertently compromise continuity, contribute to premature discharge and increase longer-term costs through relapse, readmissions and repeated care episodes.

The proposed stepped-care model offers an alternative by aligning treatment intensity, duration and continuity with measurable patient needs. Earlier intervention, appropriate alternatives to inpatient care and coordinated interdisciplinary follow-up may reduce avoidable admissions and readmissions while supporting more sustained recovery. Rather than reducing care to contain costs, the aim is to improve clinical appropriateness and achieve better value across the patient journey.

For medical schemes, this presents an opportunity to shift from managing isolated episodes and bed days towards funding coordinated, outcomes-informed care pathways. Such an approach could strengthen accountability, reduce duplication and unnecessary expenditure, and support both clinical quality and financial sustainability. Its potential clinical and economic benefits will require further evaluation.

LIMITATIONS

Several limitations should be acknowledged. The retrospective use of routinely collected clinical outcomes data limits the ability to establish causal relationships, and the reliance on registry data may introduce variability in data completeness. Similarly, the lack of inferential statistics limits the strength of the conclusions from our outcomes monitoring analysis. Future outcomes work and/or research should focus on deepening the statistical analyses of the data and the management of confounding variables.

Follow-up data were available for only a subset of patients (n=270), which may introduce selection bias in the analysis of post-discharge trajectories. Additionally, the report reflects the practices of a single provider network, which may limit generalisability to other settings.

In addition, the authors acknowledge the potential conflict of interest as two of the authors are employed by the provider network. However, the use of an independent statistician for data analysis and the inclusion of an independent, experienced third author aimed to moderate any inherent bias or vested interests. The final interpretations were sent for verification by the statistician. Corrections were made and subsequently approved. Despite these limitations, the consistency of patterns across a large sample lends support to the findings.

Finally, the proposed stepped-care model has not been tested and should therefore be understood as a data-informed framework for redesign rather than a final or universally validated solution.

FUTURE RESEARCH

Future research should focus on prospective implementation of stepped care models, including real-world cost analysis and outcome evaluation. Pilot studies of the stepped care model should be implemented across different settings in collaboration with medical schemes and funders. Future research should expand the work on prognostic modifiers, functioning, and post-discharge vulnerability, which may further refine stratification models and help identify where the current system is most vulnerable to inefficiency, duplication, and avoidable readmission. The use of inferential data analysis to support hypothesis testing, examine predictive relationships and assess the generalisability of findings should also be considered to further enhance methodological rigour in future research. Should future investigations constitute health research, it is to be submitted for review and approval by an accredited Health Research Ethics Committee before commencement, in accordance with applicable ethical and regulatory requirements.

CONCLUSION

These clinical observations highlight the limitations of diagnosis alone in understanding patient needs and recovery trajectories. While ICD-10 classification remains essential, incorporating measurable distress, functioning and other prognostic indicators can strengthen clinical decision-making and support more individualised, coordinated care through the interdisciplinary team (IDT).

A psychiatrist-led stepped-care model offers a potential pathway to align treatment intensity and continuity with patient needs, improving clinical appropriateness while creating opportunities to reduce avoidable utilisation and manage costs more effectively.

The opportunity now is for healthcare providers, clinicians and medical schemes to move beyond managing isolated services and bed days, and collaboratively develop and evaluate integrated, outcomes-informed care pathways that deliver better patient care and sustainable value.

REFERENCE LIST

Bower, P., & Gilbody, S. (2005). Stepped care in psychological therapies: Access, effectiveness and efficiency. Narrative literature review. *British Journal of Psychiatry*, *186*(1), 11–17. https://doi.org/10.1192/bjp.186.1.11

Council for Medical Schemes. (2021). *Annual report 2020/2021*. https://www.medicalschemes.co.za/cms-annual-report-2020-21/

Dalgleish, T., Black, M., Johnston, D., & Bevan, A. (2020). Transdiagnostic approaches to mental health problems: Current status and future directions. *Journal of Consulting and Clinical Psychology*, *88*(3), 179–195. https://doi.org/10.1037/ccp0000482

Drapeau, A., Marchand, A., & Beaulieu-Prévost, D. (2012). Epidemiology of psychological distress. In L. L'Abate (Ed.), *Mental illnesses: Understanding, prediction and control* (pp. 105–134). InTech. https://doi.org/10.5772/30872

Dreyer, Z., Henn, C., & Hill, C. (2019). Validation of the Depression Anxiety Stress Scale-21 (DASS-21) in a non-clinical sample of South African working adults. *Journal of Psychology in Africa*, *29*(4), 346–353. https://doi.org/10.1080/14330237.2019.1647499

Firfirey-Brijlal, N., Janse van Rensburg, A., Davis-Ferguson, M., Pegram, E., Goliath, C., Roelofse, M., Cloete, K., Goeiman, H., Smit, S., & Sulaiman-Ryklief, W. (2025). A framework for the expansion of psychosocial rehabilitation and support services in the Western Cape: Towards occupational justice and societal well-being. *South African Health Review*, *27*, 108–118. https://doi.org/10.61473/001c.141836

Fortney, J. C., Unützer, J., Wrenn, G., Pyne, J. M., Smith, G. R., Schoenbaum, M., & Harbin, H. T. (2017). A tipping point for measurement-based care. *Psychiatric Services*, *68*(2), 179–188. https://doi.org/10.1176/appi.ps.201500439

Gamieldien, F., Galvaan, R., & Silaule, O. (2025). Strengthening community-based recovery-orientated mental health services: Stakeholder insights on recovery from severe mental health conditions. *South African Health Review*, *27*, 98–107. https://doi.org/10.61473/001c.141584

Harvey, A. G., Watkins, E., Mansell, W., & Shafran, R. (2004). *Cognitive behavioural processes across psychological disorders: A transdiagnostic approach to research and treatment*. Oxford University Press. https://doi.org/10.1093/med:psych/9780198528883.001.0001

Health Quality Ontario. (2020). *Transitions between hospital and home: Care for people of all ages* [Quality standard]. https://www.hqontario.ca/Portals/0/documents/evidence/quality-standards/qs-transitions-between-hospital-and-home-quality-standard-en.pdf

Insel, T. R. (2022). *Healing: Our path from mental illness to mental health*. Penguin Press.

Lee, J., Lee, E. H., & Moon, S. H. (2019). Systematic review of the measurement properties of the Depression Anxiety Stress Scales-21 by applying updated COSMIN methodology. *Quality of Life Research*, *28*(9), 2325–2339. https://doi.org/10.1007/s11136-019-02177-x

Lewis, C. C., Boyd, M., Puspitasari, A., Navarro, E., Howard, J., Kassab, H., Hoffman, M., Scott, K., Lyon, A., Douglas, S., Simon, G., & Kroenke, K. (2019). Implementing measurement-based care in behavioral health: A review. *JAMA Psychiatry*, *76*(3), 324–335. https://doi.org/10.1001/jamapsychiatry.2018.3329

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scales* (2nd ed.). Psychology Foundation.

Mareya, S., Watts, M. C., Zhao, L., & Olasoji, M. (2024). Exploring the stepped care model in delivering primary mental health services: A scoping review. *International Journal of Mental Health Nursing*, *33*(6), 2026–2042. https://doi.org/10.1111/inm.13427

Mirowsky, J., & Ross, C. E. (2003). *Social causes of psychological distress* (2nd ed.). Aldine de Gruyter.

Mondia Health. (2025). *Celebrating sustained recovery and the value of distress levels as predictors of mental health outcomes: The Mondia Health second report (2022–2025)* [Unpublished internal report].

Mughal, S., Salmon, A., Churchill, A., Tee, K., Jaouich, A., & Shah, J. (2023). Guiding principles for implementing stepped care in mental health: Alignment on the bigger picture. *Community Mental Health Journal*, *59*(6), 1035–1042. https://doi.org/10.1007/s10597-023-01116-y

Organisation for Economic Co-operation and Development. (2021). *A new benchmark for mental health systems: Tackling the social and economic costs of mental ill-health*. OECD Health Policy Studies. OECD Publishing. https://doi.org/10.1787/4ed890f6-en

Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P. Y., Cooper, J. L., Eaton, J., Herrman, H., Herzallah, M. M., Huang, Y., Jordans, M. J. D., Kleinman, A., Medina-Mora, M. E., Morgan, E., Niaz, U., Omigbodun, O., . . . Unützer, J. (2018). The Lancet Commission on global mental health and sustainable development. *The Lancet*, *392*(10157), 1553–1598. https://doi.org/10.1016/S0140-6736(18)31612-X

Republic of South Africa. (1999). *Regulations in terms of the Medical Schemes Act, 1998 (Act No. 131 of 1998), Annexure A: Prescribed minimum benefits* (GNR 1262, Government Gazette No. 20556, 20 October 1999). Government Printer.

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. World Health Organization. https://www.who.int/publications/i/item/9241544228

World Health Organization. (2019). *International classification of diseases for mortality and morbidity statistics* (11th rev.). https://icd.who.int

World Health Organization. (2022). *World mental health report: Transforming mental health for all*. https://www.who.int/publications/i/item/9789240049338

DECLARATIONS

Acknowledgements

The authors would like to acknowledge the contribution of Marietjie Vosloo, Independent Consulting Statistician, South Africa (ORCID: 0000-0002-2041-3940), for conducting the independent statistical analyses and providing statistical consultation to support the methodological rigour of this outcomes monitoring project.

The authors further acknowledge Celesté Louw, Head: Risk and Compliance, Haven Health Management, for her contribution to systems design, governance, and compliance processes that supported the integrity and accessibility of institutional data used in the analysis.

The authors gratefully acknowledge the clinicians, administrative personnel, and operational staff across the Mondia Health network whose diligent and consistent collection of routine clinical assessment data made this outcomes monitoring report possible. Their commitment to high-quality clinical documentation and service delivery contributed substantially to institutional insights gained. Importantly, the authors acknowledge the patients whose de-identified, aggregated routine clinical data informed this report. Through their engagement with mental healthcare services, they contributed to a deeper understanding of recovery trajectories and to the development of more responsive, patient-centred systems of care.

Author contributions

Marisa Henderson: CEO, Mondia Health, South Africa. Email: marisah@mondiahealth.co.za. ORCID ID: 0009-0008-2506-4170

Juanita Swanepoel: Systems Owner: Clinical Innovation, Research and Ethics, Mondia Health, South Africa. Email: Juanitas@mondiahealth.co.za. ORCID ID: 0000-0003-4805-2312

Annamarie van Jaarsveld: Retiree, Department of Occupational Therapy, University of the Free State, South Africa. Email: annamarievanj@gmail.com. ORCID ID: 0000-0002-2357-0888

All authors have read and approved the final version of this manuscript.

Data availability statement

The de-identified, aggregated data supporting this service evaluation may be made available upon reasonable request, subject to institutional authorisation, applicable consent conditions and data-protection requirements.

Declaration of competing interest

The first two authors are employed by Mondia Health, the hospital network from which the data were obtained and which could deliver the proposed contracting model (Marisa Henderson, Chief Executive Officer; Juanita Swanepoel, Systems Owner: Clinical Innovation, Research and Ethics); the network facilitated access to patient records and institutional data. This relationship is acknowledged as a potential source of perceived bias, given the possibility that findings may be interpreted as favourable to the institution's practices, outcomes, or reputation. To mitigate this risk, a valid and reliable measurement instrument was used, and all statistical analyses were conducted independently by an external statistician using transparent analytic procedures. All three authors are occupational therapists and registered with the Health Professions Council of South Africa. Other than the institutional affiliations and potential competing interests disclosed above, the authors declare no additional competing interests.

Declaration of ethics

The authors declare that the conduct of this service evaluation and outcomes monitoring, and the writing of this manuscript adhere to ethical standards of publication, in accordance with scientific principles, and are free from plagiarism.

The authors declare that this manuscript reports on routine outcomes monitoring undertaken as part of Mondia Health's clinical outcomes and service-improvement processes. The work was not designed or conducted as research, and no additional interventions, data collection or statistical analyses were undertaken specifically for publication. Patients consented to the collection and appropriate use of routinely collected clinical information through established admission procedures. Data were analysed in de-identified form, and findings are reported in aggregate to protect patient confidentiality. All data handling was subject to applicable institutional governance procedures and the Protection of Personal Information Act 4 of 2013 (POPIA).

The authors affirm their responsibility for the accuracy, integrity and ethical reporting of the evaluation.

Declaration of assistive technologies in the writing process

The authors declare that generative artificial intelligence (GenAI) and AI-assisted tools were used in a limited and responsible manner during the preparation of this manuscript. Specifically, ChatGPT was used to support language refinement, structural organisation, clarity of expression, and editorial assistance during manuscript development. Generative AI was also used to design, revise and simplify the conceptual illustration presented in Figure 8. The authors reviewed the revised illustration for conceptual accuracy and consistency with the proposed service model. All AI-assisted content was critically reviewed, revised, and validated by the authors to ensure accuracy, originality, and compliance with ethical and scholarly standards. AI tools were not used for data collection, statistical analysis, interpretation of findings, or the generation of findings. The authors take full responsibility for the final content of this manuscript.



STREAM III

VALUE-BASED CONTRACTING AND PRICING REFORM

A COMPOSITE MARKET DISTORTION INDEX FOR REGULATORY SURVEILLANCE: DETECTING PRICE–VOLUME ANOMALIES IN GENERIC PHARMACEUTICAL MARKETS IN SOUTH AFRICA

Molaudi Gaula, Alex van den Heever, Ntobeko Mpanza, Neo Kitsa, Yongama Njisane, Selaelo Mametja, Mthokozisi Khoza, Mncengeli Sibanda, Motale Madigoe and Stanley Moloabi

How to cite this article: Gaula, M., van den Heever, A., Mpanza, N., Kitsa, N., Njisane, Y., Mametja, S., Khoza, M., Sibanda, M., Madigoe, M., & Moloabi, S. (2026). A composite market distortion index for regulatory surveillance: Detecting price–volume anomalies in generic pharmaceutical markets in South Africa. *Council for Medical Schemes Research Journal*, *1*(1), 104–120.

ABSTRACT

Background: Generic medicines markets in low- and middle-income countries (LMICs), including South Africa, are prone to price–volume distortions that persist despite regulatory frameworks such as mandatory generic substitution, the Single Exit Price (SEP) and scheme-level reimbursement policies intended to promote fair pricing and equitable access. Routine tools for detecting subtle but systematic anomalies remain limited, especially in off-patent markets. These distortions are often driven by non-transparent supplier practices, provider behaviour misaligned with policy goals, and information asymmetries that undermine the expected relationship between lower prices and higher uptake. This study therefore proposes a Composite Market Distortion Index (CMDI) to identify these distortions, enable regulatory surveillance, enhance formulary compliance and guide scheme-level oversight of medicines benefit management.

Methods: To construct the CMDI, we integrated utilisation concentration and price dispersion using the Herfindahl–Hirschman Index (HHI) and the coefficient of variation in product prices (COVPP), respectively. The CMDI was computed using a weighted geometric mean approach to flag potential price–volume anomalies. The tool was applied to 21 purposively selected off-patent product markets using two independent datasets: IQVIA national, all-payer commercial sales data (September 2014 to October 2015) and claims data from a single large medical scheme (2020–2024), thereby supporting descriptive comparisons across data sources and periods.

Results: The CMDI delineated product markets with potentially anomalous price–volume dynamics, consistent with its intended role as a screening tool. High CMDI classifications were observed in 57.14% of IQVIA product markets and 52.38% of scheme product markets, indicating broadly similar proportions of markets with elevated composite screening scores across the two datasets. High price dispersion (COVPP > 0.50) was more prevalent in the IQVIA dataset (61.90%) than in the scheme dataset (38.10%), while moderate-to-high utilisation concentration (HHI > 0.15) was common in both datasets (85.71% and 90.48%, respectively).

Conclusions: The CMDI offers a transparent, replicable tool for screening for price–volume anomalies in generic medicine markets using routinely available data. While not diagnostic of causality or misconduct, the CMDI enables regulators and funders to prioritise specific product markets for further review, assess alignment with formularies and inform targeted policy interventions, thereby strengthening pharmaceutical market surveillance in resource-constrained settings.

KEYWORDS

Generic medicines; pharmaceutical pricing; market concentration; price dispersion; composite indicators; regulatory analytics; health systems policy; South Africa

INTRODUCTION

Pharmaceutical markets in low- and middle-income countries (LMICs) are often susceptible to price–volume distortions caused by provider behaviour that does not align with policy objectives or by non-transparent supplier practices. These include financial inducements, volume-based incentives, and aggressive marketing tactics that influence prescribing and substitution choices, undermine formulary adherence, affect supplier contracting, and restrict access to affordable medicines (Gagnon, 2013; Nguyen et al., 2015; Transparency International, 2016; World Health Organization, 2020). Evidence indicates that such practices, from promotional incentives to more discreet arrangements such as kickbacks, can distort clinical decision-making and hinder the utilisation of lower-cost generics, even when they are widely available (N. K. Choudhry et al., 2002; Roy et al., 2007).

Corruption and undue influence within the pharmaceutical value chain further worsen these inefficiencies, leading to higher healthcare costs, restricted access to essential medicines, and poorer health outcomes (Transparency International, 2016). These difficulties are often compounded by information asymmetries, weak regulatory enforcement, and complex stakeholder relationships, especially in LMIC settings. Consequently, pharmaceutical pricing policies alone are frequently inadequate to guarantee efficient and equitable market outcomes (Bashaar et al., 2017).

In South Africa, significant regulatory reforms were implemented in the early 2000s to enhance transparency in pricing and guarantee fair access to medicines (National Department of Health, 1996). Central to these reforms was the introduction of the Single Exit Price (SEP) system, which set a standard ex-manufacturer price across all private-sector distribution channels and explicitly prohibited discounts, rebates, and other forms of concealed incentives (Department of Health, 2004; Williams, 2007). These measures were further supported by mandatory generic substitution policies and broader legislative efforts to regulate the distribution and dispensing margins on medicines costs (Department of Health, 2010; van den Heever, 2012).

Despite this comprehensive regulatory framework, empirical evidence shows ongoing variation in both medicine prices and utilisation patterns across clinically equivalent products. For example, studies in the South African private sector reveal significant price differences and a continued preference for higher-cost originator brands, even where lower-cost generics are available (Van der Merwe Smit & Bredenkamp, 2013). Further evidence highlights structural market features, such as pseudo-generics, products marketed by originator companies to compete with true generics, which can sustain higher prices and weaken effective competition by reducing incentives for independent generic entry (Bangalee & Suleman, 2019; Kong & Seldon, 2004; Rodrigues et al., 2014).

For their part, private medical schemes, which are central to healthcare financing in South Africa, have adopted demand-side measures such as provider networks, tiered formularies, negotiated reimbursement rates, and patient co-payments to control pharmaceutical spend (Council for Medical Schemes, 2017). However, evidence indicates that these measures do not always fully address underlying market distortions.

In the South African private sector, formulary decision-making is often limited by scarce pharmacoeconomic data and variation in evaluation processes, which may weaken the effectiveness of cost-containment strategies (Perumal-Pillay & Suleman, 2020). Moreover, co-payments can be inconsistent and affected by

factors such as medicine availability and dispensing practices, potentially restricting their intended role in promoting cost-effective utilisation. International evidence further suggests that cost-sharing mechanisms can influence utilisation but may not entirely resolve price distortions where supply-side incentives and pricing structures remain unclear (Herr & Suppliet, 2017; Vogler et al., 2013).

This gap between regulatory frameworks and market behaviour has renewed focus on pharmaceutical pricing as a key part of broader health system reform. South Africa's National Health Insurance (NHI) policy framework explicitly highlights strengthened pricing oversight and improved reimbursement design as essential for reaching universal health coverage (Department of Health, 2017).

From a theoretical perspective, well-functioning pharmaceutical markets should favour increased uptake of lower-cost, therapeutically equivalent medicines following patent expiry (Frank & Salkever, 1997; Grabowski & Vernon, 1992; Wouters et al., 2017). However, international evidence suggests that higher-priced originator medicines may retain substantial market share following generic entry despite the availability of lower-priced alternatives (Grabowski & Vernon, 1992; Van der Merwe Smit & Bredenkamp, 2013). This pattern is consistent with the market-segmentation explanation of the generic competition paradox, whereby originator prices may remain high or increase despite generic competition (Frank & Salkever, 1991, 1997; Grabowski & Vernon, 1992). This phenomenon is driven by a combination of prescriber behaviour, brand loyalty, supply-side incentives, and structural features of reimbursement systems, which can weaken price competition and limit the effectiveness of regulatory interventions (Dunne & Dunne, 2015; Dylst et al., 2013; Grabowski & Vernon, 1992; Wouters et al., 2017).

OECD evidence from oncology markets shows that negotiated rebates and reimbursement arrangements can influence actual prices paid, which may differ from published list prices (Organisation for Economic Co-operation and Development [OECD], 2020). A comparative analysis of OECD countries emphasises that pricing outcomes are affected not only by ex-manufacturer pricing policies but also by distribution margins, confidential discounts, and reimbursement structures, all of which can distort price signals and influence utilisation patterns (OECD, 2018, 2020).

Given these challenges, structured and scalable surveillance tools are needed to detect price–volume misalignments using routinely available data, thereby supporting price regulation and cost-management interventions. This study addresses this gap by introducing a Composite Market Distortion Index (CMDI) to identify anomalies in which expected relationships between price and utilisation are not observed, particularly in off-patent markets with competing alternatives. The study, therefore, aims to develop and apply a low-cost, replicable index to support formulary compliance monitoring, pricing oversight, and broader pharmaceutical policy reform within South Africa's evolving health system.

METHODS

This study employed a retrospective, observational, quantitative analytical design using routinely available secondary sales and claims data to identify potential price–volume distortions in off-patent pharmaceutical markets in South Africa. The approach is founded on established principles of market analysis and composite indicator development, which facilitate the use of quantitative metrics to simplify multidimensional phenomena into understandable signals for policy and regulation (Greco et al., 2019; OECD, 2008). Composite indicators are particularly valuable when complex systems cannot be effectively described by single measures, enabling the integration of multiple dimensions into a single analytical model (Greco et al., 2019; OECD, 2008).

Product markets definition and product selection

Product-specific markets were defined using the Anatomical Therapeutic Chemical (ATC) classification system and aligned with South Africa's regulatory framework for interchangeable multi-source medicines (generics). In accordance with the Medicines and Related Substances Act (Act 101 of 1965, as amended), generic medicines are defined as products that are pharmaceutically equivalent and bioequivalent to a reference medicine and are eligible for substitution under specified conditions. Based on this framework, each product market comprised interchangeable medicines sharing the same active pharmaceutical ingredient (API), dosage form, route of administration, and strength, and limited to off-patent products. This approach to selecting product markets for analysis is consistent with the well-accepted principle of demand- and supply-side substitution, which is used to define relevant markets for assessing competitive interactions between homogeneous products (Bishop & Walker, 2010). Using the ATC criteria and the definition of multi-source medicines, we purposively selected and defined 21 off-patent product markets for this analysis.

Data sources and analytical approach

The CMDI was developed and applied to two distinct datasets: (i) national commercial pharmaceutical sales data obtained from IQVIA (2014–2015), representing an all-payer view of medicine sales across the South African private market, and (ii) funder-level pharmacy claims data (2020–2024) from a large South African medical scheme, representing reimbursed utilisation within a single managed healthcare financing environment. These datasets offered product-specific details on volumes and prices, facilitating consistent calculation of market indicators across different contexts, datasets, and timeframes. The two datasets therefore provide complementary perspectives on pharmaceutical market behaviour, with the IQVIA dataset reflecting broader market activity and the scheme dataset reflecting utilisation patterns within a specific reimbursement setting. The utilisation of multiple data sources enhances robustness and external validity, aligning with best practices in composite indicator development (OECD, 2008).

Analytical assumptions

The analytical framework is grounded in economic and behavioural theory, particularly the rational economic agent model, which assumes that decision-makers seek to maximise utility under conditions of full information when choosing among therapeutically equivalent options (Arrow, 1978; Marschak, 1950). Under these conditions, the model predicts a consistent preference for lower-cost options, as higher-priced alternatives offer no additional value. In this context, prescribers, dispensers, and payers are likely to favour lower-priced medicines when clinical outcomes are comparable, thereby increasing market share for more affordable products. However, pharmaceutical markets involve complex interactions among manufacturers, prescribers, and payers, in which financial incentives, behavioural biases, and information gaps can distort decision-making and undermine the expected relationship between price and utilisation (Manchanda et al., 2005). These variations provide the basis for recognising anomalies in price–volume relationships.

COMPUTATION OF PRODUCT MARKET INDICATORS

Market concentration: Herfindahl–Hirschman Index

The Herfindahl–Hirschman Index (HHI) was employed as a signalling indicator to assess the distribution of utilisation, based on sales volumes, across therapeutically equivalent products within defined generic markets (Bishop & Walker, 2010). Although it is commonly used in competition economics to evaluate market concentration and its related implications for market power or competition between firms, in this study, the HHI is utilised descriptively to characterise the degree of concentration or dispersion in product volumes (Motta, 2004). A key aspect of the HHI relevant to assessing dispersion in product volumes is that it measures an individual product's share of total market volumes and, in doing so, accounts for the relative size

distribution of products within the specified product market (Cracau & Lima, 2016; Matsumoto et al., 2012). The reported n is the number of analytical product observations in each product market, with separate pack-size presentations retained as distinct observations. Mean market share (MMS) and HHI were calculated from the same observations; where volume shares sum to one, MMS = 1/n. HHI was calculated as the sum of squared observation-level market shares:

$$HHI = \sum_{i=1}^{n} S_i^2$$

Where S_i represents the unit volume-based market share of the product i within the defined market. HHI values range from near 0, indicating a more even distribution of utilisation across products, to near 1, indicating greater concentration of volumes among fewer products (Cracau & Lima, 2016). Higher HHI values in this context may signal uneven utilisation patterns across therapeutically equivalent products and are interpreted as indicative of potential anomalies in volume distribution.

Price variations: coefficient of variation

Intra-market price dispersion was assessed using the coefficient of variation among individual product prices (COVPP). For IQVIA, product-level mean values were calculated by dividing Total Value by Total Units. For scheme claims, presentation-level mean claimed amounts were calculated as total claimed expenditure divided by claim count. No separate VAT adjustment was applied; the analysis used the expenditure/value fields as supplied in the respective source datasets. Within each product market, the mean market price (MMP) was calculated as the unweighted arithmetic mean of these observation-level values, and COVPP as their population standard deviation divided by their mean:

$$COVPP = \sigma_p / \mu_p$$

Where σ_p denotes the population standard deviation of the observation-level values and μ_p denotes their arithmetic mean. Cross-dataset comparisons were interpreted descriptively, recognising potential differences in source-recording conventions and pack-size composition. In well-functioning markets for clinically interchangeable medicines, prices are expected to converge as lower-priced products capture a larger share of volumes. The COVPP is a standard statistical measure of relative variability and is widely used to assess dispersion across comparable units because it expresses variability relative to the mean, facilitating comparisons across markets with different absolute price levels (Levine et al., 2008). In pharmaceutical markets, lower dispersion is expected in well-functioning generic markets due to price competition and convergence towards lower-priced alternatives, whereas higher dispersion may indicate inefficiencies, weak substitution dynamics, or persistent segmentation among clinically equivalent products (Atal et al., 2016; Galizzi et al., 2011).

Composite Market Distortion Index

To integrate price and volume dynamics into a unified screening tool, the CMDI was constructed by combining HHI and COVPP using a weighted geometric mean:

$$CMDI = HHI^{\alpha} \times COVPP^{(1-\alpha)}$$

A weighting of $\alpha = 0.4$ was assigned to HHI and $(1 - \alpha) = 0.6$ to COVPP, reflecting a theory-informed emphasis on price variation as the more direct manifestation of potential distortion in generic pharmaceutical markets, while retaining substantial influence from HHI as a behavioural proxy for utilisation. In well-functioning markets, clinically equivalent generics would be expected to show limited price dispersion as competition encourages price convergence; persistent price variation among therapeutically interchangeable

products may therefore signal market inefficiencies, information asymmetries, prescribing and dispensing behaviours, reimbursement dynamics or other structural distortions. Market concentration indicates how utilisation is distributed across clinically similar products relative to prevailing prices. The CMDI was designed primarily as a screening tool to identify product markets in which unexpected price–volume relationships may warrant further investigation. The use of geometric aggregation is consistent with composite indicator methodology, as it reduces compensability between components and preserves proportional relationships (Greco et al., 2019; OECD, 2008). Weighting decisions were informed by methodological guidance, recognising that weighting and aggregation are critical determinants of composite indicator performance and interpretation (Greco et al., 2019; Perišić, 2015). The CMDI was developed as an exploratory screening tool rather than a prescriptive regulatory measure, and the selected weights should therefore be interpreted as illustrative rather than definitive, consistent with recommendations that composite indicator assumptions be made transparent and interpreted cautiously (Perišić, 2015; Santeramo, 2017).

No formal sensitivity analysis of the alternative weighting structures was undertaken in the present study. Future research should assess the robustness of CMDI classifications under alternative weighting assumptions and evaluate the influence of weighting choices on market rankings and distortion classifications, as sensitivity and robustness testing are recognised as essential components of composite indicator development (Greco et al., 2019; OECD, 2008; Perišić, 2015). For this analysis, a benchmark threshold of CMDI = 0.23 was derived from conceptual screening reference values of HHI = 0.15 and COVPP = 0.30. The HHI reference value of 0.15 was informed by the historical 2010 DOJ/FTC Horizontal Merger Guidelines, which were superseded in 2023. The study's observation-level HHI categories are exploratory screening ranges and do not reproduce regulatory merger-assessment thresholds (Motta, 2004; U.S. Department of Justice & Federal Trade Commission, 2010). The COVPP benchmark of 0.30 was selected as a pragmatic reference value representing relatively low price dispersion among clinically interchangeable generic medicines. As no universally accepted threshold exists for price dispersion in pharmaceutical markets, this value was specified a priori to represent a market in which prices exhibit limited variability and are broadly consistent with expectations of effective generic competition. These reference values were specified a priori, before analysis, and used to define a conceptual benchmark for expected market functioning rather than regulatory or empirically validated cut-offs.

Applying the selected weighting structure (α = 0.4) to these reference values yielded CMDI = $0.15^{0.4} \times 0.30^{0.6} = 0.227$, rounded to approximately 0.23, which was used as a screening threshold to identify product markets warranting further investigation. Threshold categories for HHI, COVPP and CMDI (Low, Moderate and High) were similarly established as interpretive screening ranges to support prioritisation and comparison across markets and should not be interpreted as definitive indicators of market failure or regulatory non-compliance. Product markets exceeding this threshold were flagged as potentially distorted and prioritised for further review.

Rationale for indicator selection

The selection of HHI and COVPP reflects their complementary ability to capture volume concentration and price dispersion, respectively. Together, these indicators provide a parsimonious representation of market dynamics using routinely available data. Their integration into a composite index aligns with established approaches in policy analytics, which combine multiple indicators to enhance interpretability while maintaining transparency (OECD, 2008; Santeramo, 2017).

RESULTS

Across the 21 off-patent product markets assessed, variation was observed in both price dispersion and utilisation concentration across the IQVIA and scheme datasets (Table 1). Mean HHI values were broadly

similar between IQVIA and scheme claims (0.41 and 0.39, respectively), while mean COVPP was 0.67 in IQVIA and 0.61 in scheme claims. Mean CMDI values were 0.51 and 0.42, respectively.

Using the predefined screening ranges, 57.14% of IQVIA markets (12/21) and 52.38% of scheme markets (11/21) were classified as High CMDI, while 33.33% of markets in each dataset were classified as Moderate CMDI. Low CMDI classifications accounted for 9.52% of IQVIA markets and 14.29% of scheme markets (Table 2). High price dispersion (COVPP > 0.50) was more prevalent in IQVIA than in scheme claims (61.90% versus 38.10%), whereas Moderate-to-High HHI classifications (> 0.15) were common in both datasets (85.71% and 90.48%, respectively). These results indicate broadly similar CMDI classification profiles despite differences in the distribution of the component indicators.

At product-market level, the highest IQVIA CMDI values were observed for Acetylsalicylic Acid 300 mg (0.95), Paracetamol 500 mg (0.91), Aluminium Hydroxide 50 mg/ml (0.80), Fluoxetine 20 mg (0.76) and Doxycycline 100 mg (0.71). In scheme claims, the highest CMDI values were observed for Paracetamol 500 mg (0.90), Fluoxetine 20 mg (0.77), Acetylsalicylic Acid 300 mg (0.69), Erythromycin 250 mg (0.66) and Ceftriaxone 1,000 mg (0.59).

Several markets showed similar CMDI classifications across the two datasets. For example, Paracetamol 500 mg had CMDI values of 0.91 in IQVIA and 0.90 in scheme claims; Fluoxetine 20 mg had values of 0.76 and 0.77; and Ceftriaxone 1,000 mg had values of 0.64 and 0.59, respectively. Cefuroxime Axetil 250 mg remained in the Low CMDI range in both datasets (0.10 and 0.16). Conversely, differences were evident in some markets: Temozolomide 120 mg had a CMDI of 0.43 in IQVIA and 0.04 in scheme claims, while Escitalopram 10 mg had a Moderate CMDI score in IQVIA (0.36) and a High score in scheme claims (0.50).

Overall, the CMDI identified both convergent and divergent price–utilisation patterns across the two datasets. Differences in individual HHI and COVPP profiles were not necessarily reflected in corresponding differences in the composite classification, reinforcing the intended use of the CMDI as a descriptive screening measure for prioritising product markets for further investigation rather than as a diagnostic measure of market failure or regulatory non-compliance.

Product-level analysis

To illustrate the application of the CMDI framework, price–volume plots were generated for two contrasting IQVIA product markets (Figures 1 and 2). The plots show observation-level mean prices and volume shares relative to the corresponding mean market price (MMP) and mean market share (MMS), illustrating the distributions underlying the screening indicators.

Cefuroxime Axetil 250 mg tablets (Figure 1) showed a Low-CMDI profile. Most prices clustered around the MMP of R85.60, although volume shares varied relative to the MMS of 8.33%. The highest-share observation exceeded both means (R108.65; 19.23%). Overall, the market showed low utilisation concentration (HHI = 0.13), low price dispersion (COVPP = 0.09) and a Low CMDI score (0.10).

In contrast, Tenofovir–Emtricitabine–Efavirenz 300/200/600 mg tablets (Figure 2) showed greater price dispersion, with an MMP of R271.54 and an MMS of 11.11%. Observation M5.1 was priced at R426.82 and accounted for 50.44% of volume, placing it substantially above both means. The market showed moderate utilisation concentration (HHI = 0.30), high price dispersion (COVPP = 0.52) and a High CMDI score (0.41). The plot illustrates the underlying distribution, with substantial utilisation concentrated in a relatively high-priced observation.

Table 1 provides the detailed comparative indicator profiles for the 21 selected product markets across IQVIA and scheme claims, including n, MMP, MMS, HHI, COVPP and CMDI. Values are rounded to two decimal

places; classifications used the unrounded values.

Table 1: Comparative market indicator profiles for selected generic medicine markets: IQVIA sales data versus scheme claims data

	IQVIA market data (Sept 2014–Oct 2015)						Scheme claims data (2020–2024)					
Product	n	MMP	MMS	HHI	COVPP	CMDI	n	MMP	MMS	HHI	COVPP	CMDI
Doxycycline 100 mg capsules	7	R182.85	14.29%	0.55	0.85	0.71	3	R62.98	33.33%	0.41	0.44	0.43
Amoxicillin 500 mg capsules	16	R128.09	6.25%	0.34	0.87	0.60	20	R219.74	5.00%	0.33	0.78	0.55
Clavulanic acid 1,000 mg tablets	8	R218.90	12.50%	0.24	1.28	0.66	16	R270.64	6.25%	0.17	1.05	0.51
Cefuroxime axetil 250 mg tablets	12	R85.60	8.33%	0.13	0.09	0.10	13	R172.04	7.69%	0.21	0.14	0.16
Ceftriaxone 1,000 mg FPB vial (dry)	11	R94.10	9.09%	0.25	1.20	0.64	23	R120.21	4.35%	0.22	1.12	0.59
Erythromycin 250 mg capsules	6	R219.82	16.67%	0.42	0.77	0.60	5	R272.61	20.00%	0.37	0.97	0.66
Tenofovir–emtricitabine–efavirenz 300/200/600 mg tablets	9	R271.54	11.11%	0.30	0.52	0.41	28	R385.40	3.57%	0.13	0.34	0.23
Efavirenz 600 mg tablets	10	R335.77	10.00%	0.11	0.69	0.33	11	R225.68	9.09%	0.38	0.31	0.34
Nevirapine 200 mg tablets	10	R189.12	10.00%	0.11	0.50	0.28	10	R299.28	10.00%	0.42	0.19	0.26
Lamivudine 150 mg tablets	8	R178.76	12.50%	0.29	0.46	0.38	9	R173.63	11.11%	0.37	0.23	0.28
Tenofovir disoproxil fumarate 300 mg tablets	10	R176.52	10.00%	0.31	0.46	0.39	7	R328.26	14.29%	0.33	0.23	0.26
Escitalopram 10 mg tablets	15	R87.61	6.67%	0.28	0.42	0.36	19	R129.91	5.26%	0.54	0.48	0.50
Citalopram 20 mg tablets	17	R73.17	5.88%	0.56	0.70	0.64	19	R129.23	5.26%	0.24	0.62	0.43
Fluoxetine 20 mg capsules	14	R52.27	7.14%	0.35	1.27	0.76	16	R97.02	6.25%	0.47	1.08	0.77
Temozolomide 120 mg capsules	3	R2,103.23	33.33%	1.00	0.25	0.43	2	R2,447.42	50.00%	0.53	0.01	0.04
Gemcitabine 1 g injection	3	R1,323.29	33.33%	0.40	0.04	0.10	3	R1,572.28	33.33%	0.61	0.21	0.33
Paclitaxel 30 mg infusion	4	R215.65	25.00%	0.44	0.23	0.29	7	R167.91	14.29%	0.46	0.03	0.09
Letrozole 2.5 mg tablets	3	R673.50	33.33%	0.41	0.35	0.37	7	R654.05	14.29%	0.42	0.37	0.39
Paracetamol 500 mg tablets	12	R15.87	8.33%	0.52	1.31	0.91	54	R89.22	1.85%	0.14	3.05	0.90
Acetylsalicylic acid 300 mg tablets	7	R28.23	14.29%	0.76	1.09	0.95	10	R36.38	10.00%	0.62	0.74	0.69
Aluminium hydroxide 50 mg/ml suspension	9	R39.65	11.11%	0.91	0.73	0.80	2	R129.31	50.00%	0.79	0.36	0.49
Average	**9.24**	**R318.74**	**14.25%**	**0.41**	**0.67**	**0.51**	**13.52**	**R380.15**	**15.01%**	**0.39**	**0.61**	**0.42**
Standard deviation	**4.08**	**R488.30**	**8.82%**	**0.23**	**0.39**	**0.24**	**11.53**	**R561.58**	**14.08%**	**0.17**	**0.64**	**0.22**
Minimum	**3**	**R15.87**	**5.88%**	**0.11**	**0.04**	**0.10**	**2**	**R36.38**	**1.85%**	**0.13**	**0.01**	**0.04**
Maximum	**17**	**R2,103.23**	**33.33%**	**1.00**	**1.31**	**0.95**	**54**	**R2,447.42**	**50.00%**	**0.79**	**3.05**	**0.90**
Conceptual reference levels for screening				0.15	0.30	0.23						

Note: The table compares pricing and utilisation indicators across selected generic medicine markets using IQVIA market data (formerly IMS Health; September 2014–October 2015) and scheme claims data. Mean market price (MMP) summarises average observation-level monetary values, while mean market share (MMS) is the arithmetic mean of observation-level volume shares. The Herfindahl–Hirschman Index (HHI; 0–1) describes utilisation concentration, the coefficient of variation in product prices (COVPP) measures relative price dispersion, and the Composite Market Distortion Index (CMDI) combines HHI and COVPP to identify potential market anomalies. MMP and COVPP were calculated from total value divided by total units for IQVIA and total claimed expenditure divided by claim count for scheme claims. Cross-dataset comparisons are descriptive, recognising potential differences in source-recording conventions and pack-size composition. The reported n represents the number of analytical product observations included within each product market, with separate pack-size presentations retained as distinct observations. MMS and HHI were calculated using the same set of included analytical observations; consequently, where volume shares sum to one, MMS = 1/n.

Table 2: Distribution of market indicators by threshold category for the IQVIA and scheme datasets (N = 21 product markets per dataset)

Indicator	Threshold category	IQVIA dataset		Scheme dataset	
		n	%	n	%
HHI	High	6	28.57%	5	23.81%
	Moderate	12	57.14%	14	66.67%
	Low	3	14.29%	2	9.52%
COVPP	High	13	61.90%	8	38.10%
	Moderate	4	19.05%	6	28.57%
	Low	4	19.05%	7	33.33%
CMDI	High	12	57.14%	11	52.38%
	Moderate	7	33.33%	7	33.33%
	Low	2	9.52%	3	14.29%

Note: Thresholds used for indicator classification. HHI: low ≤ 0.15; moderate > 0.15 to ≤ 0.50; high > 0.50. COVPP: low ≤ 0.30; moderate > 0.30 to ≤ 0.50; high > 0.50. CMDI: low ≤ 0.23; moderate > 0.23 to ≤ 0.40; high > 0.40. Threshold categories represent interpretive screening ranges used to support comparison and prioritisation across product markets and should not be interpreted as definitive indicators of optimal market functioning, market failure or regulatory non-compliance. Classifications were assigned using the underlying unrounded indicator values; values presented in the tables were rounded to two decimal places. Abbreviations: CMDI = Composite Market Distortion Index; COVPP = coefficient of variation in product prices; HHI = Herfindahl–Hirschman Index.

Table 2 summarises the distribution of markets across Low, Moderate and High HHI, COVPP and CMDI screening categories for both datasets, enabling a direct comparison of the component and composite indicator distributions.

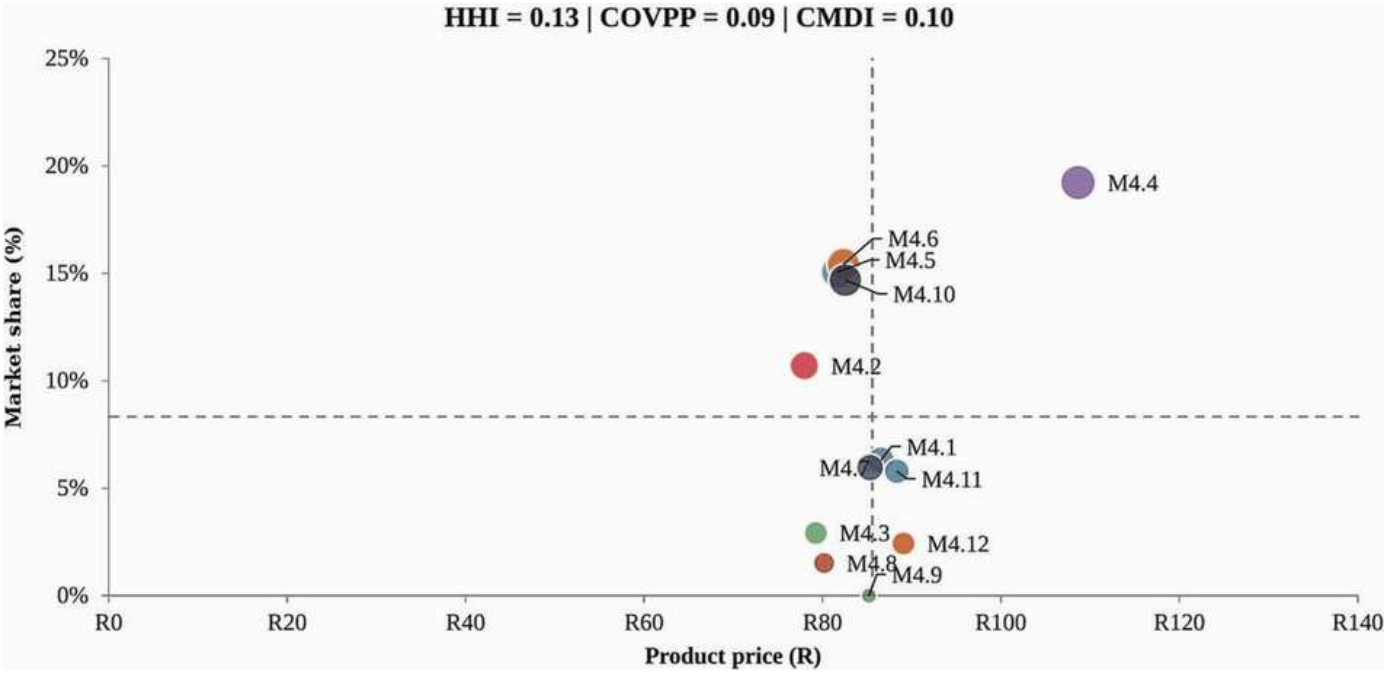

Figure 1: Price–volume alignment in a low-CMDI market for cefuroxime axetil 250 mg tablets (expected pattern)

Note: Distribution of observation-level mean prices and volume shares for cefuroxime axetil 250 mg tablets in the IQVIA dataset (September 2014–October 2015). The figure displays 12 analytical observations. Dashed lines indicate MMP = R85.60 and MMS = 8.33%. HHI = 0.13, COVPP = 0.09 and CMDI = 0.10.

Figure 1 illustrates the expected Low-CMDI price–volume pattern for Cefuroxime Axetil 250 mg tablets.

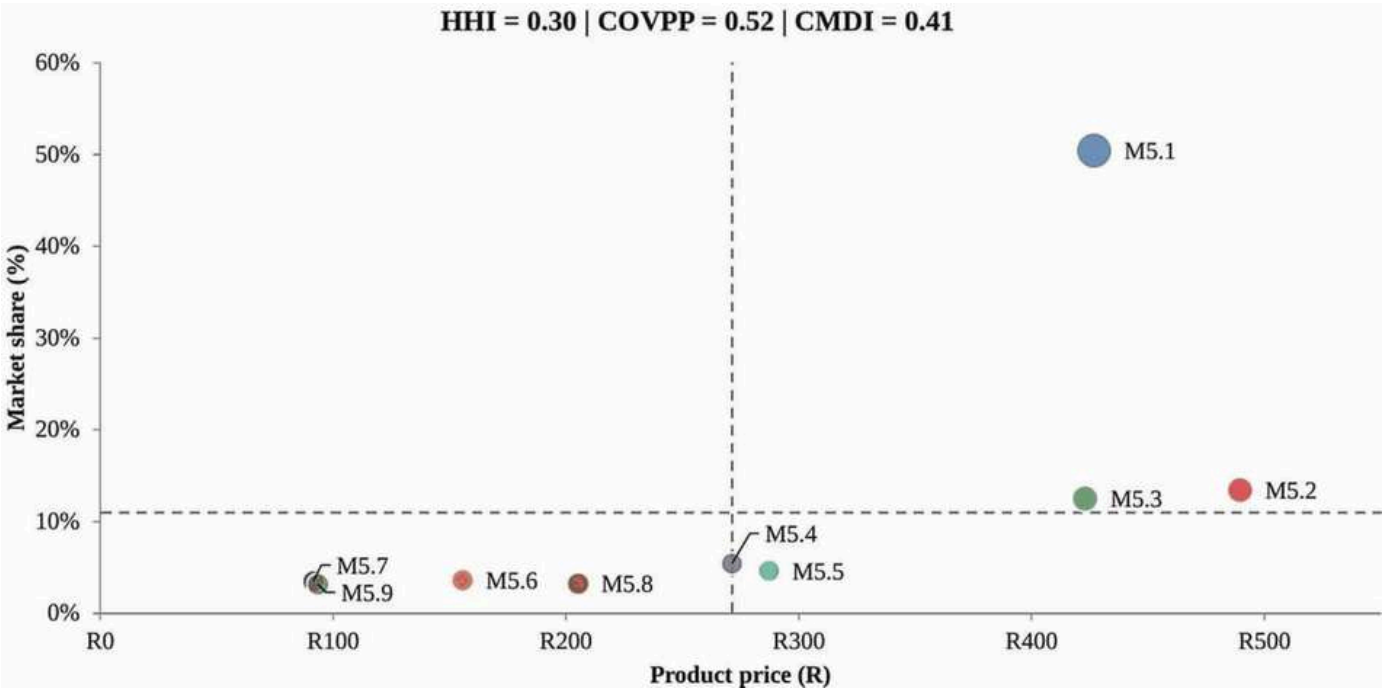

Figure 2: Price–volume misalignment in a high-CMDI market for tenofovir–emtricitabine–efavirenz 300/200/600 mg tablets (unexpected pattern)

Note: Distribution of observation-level mean prices and volume shares for tenofovir–emtricitabine–efavirenz 300/200/600 mg tablets in the IQVIA dataset (September 2014–October 2015). The figure displays nine analytical observations. Dashed lines indicate MMP = R271.54 and MMS = 11.11%; these are descriptive means rather than additional product observations. The market shows moderate utilisation concentration (HHI = 0.30), high price dispersion (COVPP = 0.52) and a high CMDI score (0.41). The distribution warrants further investigation but does not establish the factors responsible for the observed utilisation pattern.

Figure 2 shows the price–volume distribution for Tenofovir–Emtricitabine–Efavirenz 300/200/600 mg tablets, with moderate utilisation concentration (HHI = 0.30), high price dispersion (COVPP = 0.52) and a High CMDI score (0.41).

DISCUSSION

This study examined pricing and utilisation patterns across 21 off-patent generic pharmaceutical markets using IQVIA national sales data (September 2014–October 2015) and scheme claims data (2020–2024). By combining observation-level utilisation concentration and price dispersion, the CMDI identified product markets warranting further investigation in both datasets, illustrating its potential application as a screening tool for regulators and funders.

High CMDI scores (> 0.40) were observed in 57.14% of IQVIA markets and 52.38% of scheme markets. These broadly similar proportions occurred despite differences in the component indicators and should not be interpreted as evidence of equivalent market conditions. The datasets represent different periods, market coverage and financing environments, with potential differences in source-recording conventions and pack-size composition. Observed differences may therefore reflect temporal changes, reimbursement arrangements, generic uptake and data capture, as well as underlying market behaviour. Such contextual and structural differences are relevant when interpreting composite indicators (Greco et al., 2019; OECD, 2008). Application across both datasets illustrates the framework's potential adaptability, although it does not establish external validity (Kaplan & Laing, 2005; Vogler et al., 2017). International evidence similarly shows that pharmaceutical price formation varies across regulatory environments and product types (Cabrales & Jiménez-Martín, 2013).

Low utilisation concentration (HHI ≤ 0.15) was observed in 14.29% of IQVIA markets and 9.52% of scheme markets, whereas High HHI values (> 0.50) occurred in 28.57% and 23.81%, respectively. Low CMDI scores (≤ 0.23) were observed in 9.52% of IQVIA markets and 14.29% of scheme markets. These categories represent descriptive screening ranges rather than evidence of ideal market functioning. Higher HHI values indicate concentration of utilisation among fewer analytical product observations; they do not independently establish supplier market power or preferential uptake of higher-priced products. The findings provide a basis for examining concerns about the SEP framework's reach in addressing net pricing strategies and preferential contracting, without establishing that these practices explain the observed patterns (Gray & Suleman, 2015). Evidence also suggests that price regulation alone may not substantially alter utilisation where entry dynamics, therapeutic choice or other non-price factors influence market shares (Hsiao et al., 2010). The Medicines Act and national industry code provide a relevant governance context through restrictions on inducements linked to prescribing, stocking, dispensing, administering or purchasing health products (Department of Health, 2010; Marketing Code Authority of South Africa, 2014).

High price dispersion (COVPP > 0.50) was more prevalent in IQVIA markets (61.90%) than in scheme markets (38.10%). Substantial price variation among therapeutically equivalent medicines has also been reported in other regulated pharmaceutical markets, suggesting that formal price controls do not necessarily eliminate price variability (Atal et al., 2016; Shrestha et al., 2020). Although payer controls may contribute to differences between the datasets, their effects cannot be separated here from those of differences in periods, product composition and recording conventions. Nevertheless, 85.71% of scheme markets had Moderate or High CMDI scores, indicating that screening signals remained common within the managed reimbursement setting. International comparisons of generic medicine markets provide further context for variation in prices and uptake across health systems (Wouters et al., 2017). Evidence from regulated generic markets shows that policy reforms may reduce prices without uniformly reducing dominant firms' market shares (Linden, 2013). Evidence from China similarly highlights the importance of institutional incentives where price controls coexist with declining generic market share and rising weighted-average generic prices (Zeng, 2013).

International evidence identifies opaque net pricing structures, information asymmetries and prescriber or dispenser incentives as potential contributors to inefficiencies in off-patent markets (Dylst et al., 2012; Morgan et al., 2017). Persistent price differences among clinically equivalent products have also been documented in low- and middle-income countries, indicating that competition does not invariably produce price convergence (Das et al., 2007; Shrestha et al., 2020). Behavioural preferences, supplier incentives and institutional arrangements may help explain persistent segmentation in these markets, although their contribution was not assessed in this study (Godman et al., 2014; Nguyen et al., 2015).

The CMDI framework uses existing administrative or commercial data to support market surveillance. Its geometric aggregation reduces compensability between utilisation concentration and price dispersion, consistent with composite-indicator methodology (Greco et al., 2019; OECD, 2008). The illustrative plots show how these components relate to the underlying observations. Cefuroxime Axetil 250 mg tablets had low price dispersion and utilisation concentration, whereas the largest-share observation in the Tenofovir–Emtricitabine–Efavirenz market was among the higher-priced observations. The latter pattern warrants investigation but does not establish why the product attracted greater utilisation. Financial conflicts, including kickbacks and self-referral arrangements, are potential influences identified in the wider literature, rather than explanations demonstrated by these data (S. Choudhry et al., 2004; Jain et al., 2014). South Africa's self-regulatory marketing code provides a relevant framework by restricting inducements and requiring remuneration for legitimate activities to reflect fair market value (Marketing Code Authority of South Africa, 2014).

These insights are particularly relevant to South Africa's National Health Insurance Act 20 of 2023 (Republic of South Africa, 2024), which provides for strategic purchasing to support universal access to healthcare. The

CMDI could support formulary governance, procurement planning, and pricing surveillance by identifying product markets that require closer review. Realising these benefits would also require regulatory capacity, aligned incentives and ongoing monitoring of purchasing and delivery arrangements.

The study provides an exploratory approach to pharmaceutical market screening rather than a causal or independently validated assessment of market distortion. Future research should apply the CMDI to time-aligned longitudinal datasets, assess the classification sensitivity to alternative weights and screening thresholds, and examine its usefulness relative to independently reviewed market conditions. Incorporating time since loss of exclusivity, generic entry, and duration of competition could also help distinguish market-development effects from patterns that warrant further investigation.

LIMITATIONS

Prices were calculated using the source-specific denominators described in the Methods: Total Value divided by Total Units for IQVIA, and total claimed expenditure divided by claim count for scheme claims. Cross-dataset comparisons are therefore descriptive, recognising potential differences in recording conventions and pack-size composition. Separate pack-size presentations were retained, and COVPP should not be interpreted solely as variation in prices per equivalent physical unit. The indicators are also sensitive to the analytical unit: retaining separate pack-size presentations as distinct analytical observations can yield different HHI and CMDI values from aggregation at a broader product level. The scores therefore describe utilisation patterns across the included analytical product observations and should not be interpreted directly as concentration between brands or suppliers. No systematic sensitivity analysis of alternative aggregation levels was undertaken; future validation should assess their effects on scores, rankings and classifications. The analysis also did not explicitly account for market maturity, including time since patent expiry, generic entry, and duration of competition, because these variables were not consistently available. The purposive selection of 21 markets and absence of formal sensitivity analysis further limit generalisation and conclusions about the robustness of the screening classifications.

CONCLUSION

Despite South Africa's SEP policy framework, the selected off-patent product markets displayed variation in prices and utilisation concentration, highlighting patterns relevant to competitive pricing and rational medicine use. The CMDI provides a transparent, reproducible approach to screening for potential price–volume anomalies using routinely available data. Its application identified markets that warrant further review in both datasets, although comparisons remain descriptive given differences in periods, coverage and recording conventions. The index does not establish market failure, causality or misconduct, but could support regulatory oversight, policy prioritisation and purchasing reviews. Further validation and sensitivity assessment would strengthen its application to pharmaceutical market surveillance as South Africa progresses towards UHC under the NHI.

POLICY IMPLICATIONS

To support the goals of the NHI and strengthen pharmaceutical governance, price regulation should be complemented by measures addressing purchasing arrangements and utilisation. International evidence highlights the importance of transparency regarding actual transaction prices, including negotiated discounts and rebates that may not be reflected in published list prices (OECD, 2018; Vogler et al., 2013). The long-term sustainability of off-patent markets also depends on balancing supply-side price discipline with demand-side measures that stimulate cost-effective uptake; evidence from Spain suggests that continuous downward pressure on generic prices, when combined with limited diffusion, can weaken the sustainability of generic markets and reduce their competitive effect on off-patent prices (Dylst et al., 2014).

Claims data can support monitoring of prescribing and dispensing patterns and help identify potential inefficiencies requiring further investigation, including circumstances in which financial incentives may influence clinical decision-making (Nguyen et al., 2015; Wouters et al., 2017). Greater coordination of formulary and purchasing arrangements across schemes and levels of care may also support equitable access and cost-effective generic substitution (Dylst et al., 2012; Morgan et al., 2017). Evidence from both Europe and China suggests that supply-side reforms alone are insufficient; effective generic uptake typically requires coordinated demand-side measures, including prescribing guidance, pharmacist substitution, patient incentives, and trust-building interventions around generic quality and equivalence (Dylst et al., 2014; Zeng, 2013).

Strengthening routine analytical capacity could enable timely identification and review of potential market anomalies using tools such as the CMDI. Transparent indicator construction and clearly specified screening criteria are important foundations for such applications (Godman et al., 2014; OECD, 2008). In parallel, governance mechanisms prohibiting inducements, kickbacks and other pecuniary advantages linked to prescribing or dispensing decisions should be monitored and enforced, although the present study does not establish that these practices explain the observed patterns (Marketing Code Authority of South Africa, 2014). Together, these measures could strengthen pharmaceutical market oversight and support more informed purchasing and policy decisions.

REFERENCE LIST

Arrow, K. J. (1978). Uncertainty and the welfare economics of medical care. In *Uncertainty in economics* (pp. 345–375). Elsevier.

Atal, S., Atal, S., Deshmankar, B., & Nawaz, S. A. (2016). Cost analysis of commonly used drugs under price control in India: Assessing the effect of drug price control order on brand price variation. *International Journal of Pharmacy and Pharmaceutical Sciences*, *8*, 315–321.

Bangalee, V., & Suleman, F. (2019). Pseudo-generics in South Africa: A price comparison study. *Value in Health Regional Issues*, *19*, 87–91. https://doi.org/10.1016/j.vhri.2019.06.001

Bashaar, M., Hassali, M. A., Saleem, F., Alrasheedy, A. A., Thawani, V., & Babar, Z.-U.-D. (2017). Efficacy of international approaches to medicine price regulation and control: A scoping review. *Journal of Applied Pharmaceutical Science*, *7*, 227–241. https://doi.org/10.7324/JAPS.2017.70434

Bishop, S., & Walker, M. (2010). *The economics of EC competition law: Concepts, application and measurement*. Sweet & Maxwell.

Cabrales, A., & Jiménez-Martín, S. (2013). The determinants of pricing in pharmaceuticals: Are US prices really so high? *Health Economics*, *22*, 1377–1397.

Choudhry, N. K., Stelfox, H. T., & Detsky, A. S. (2002). Relationships between authors of clinical practice guidelines and the pharmaceutical industry. *JAMA*, *287*(5), 612–617. https://doi.org/10.1001/jama.287.5.612

Choudhry, S., Choudhry, N. K., & Brown, A. D. (2004). Unregulated private markets for health care in Canada? Rules of professional misconduct, physician kickbacks and physician self-referral. *CMAJ*, *170*, 1115–1118.

Council for Medical Schemes. (2017). *Annual report 2016/2017*. https://www.medicalschemes.co.za

Cracau, D., & Lima, J. E. D. (2016). On the normalized Herfindahl-Hirschman index: A technical note. *International Journal on Food System Dynamics*, *7*(4), 382–386.

Das, S. C., Mandal, M., & Mandal, S. C. (2007). A critical study on availability and price variation between different brands: Impact on access to medicines. *Indian Journal of Pharmaceutical Sciences*, *69*(1), 160–163. https://doi.org/10.4103/0250-474X.32139

Department of Health. (2004). *Regulations relating to a transparent pricing system for medicines and scheduled substances* (Government Notice No. R. 553).

Department of Health. (2010). *Medicines and Related Substances Act (Act No. 101 of 1965, as amended)*.

Department of Health. (2017). *National Health Insurance for South Africa: Towards universal health coverage* [White paper].

Dunne, S. S., & Dunne, C. P. (2015). What do people really think of generic medicines? A systematic review and critical appraisal of literature on stakeholder perceptions of generic drugs. *BMC Medicine*, *13*, Article 173.

Dylst, P., Vulto, A., & Simoens, S. (2012). Reference pricing systems in Europe: Characteristics and consequences. *Generics and Biosimilars Initiative Journal*, *1*, 127–131.

Dylst, P., Vulto, A., & Simoens, S. (2013). Demand-side policies to encourage the use of generic medicines: An overview. *Expert Review of Pharmacoeconomics & Outcomes Research*, *13*, 59–72.

Dylst, P., Vulto, A. G., & Simoens, S. (2014). Analysis of Spanish generic medicines retail market: Recommendations to enhance long-term sustainability. *Expert Review of Pharmacoeconomics & Outcomes Research*, *14*, 345–353.

Frank, R. G., & Salkever, D. S. (1991). *Pricing, patent loss and the market for pharmaceuticals* (NBER Working Paper No. 3803). National Bureau of Economic Research.

Frank, R. G., & Salkever, D. S. (1997). Generic entry and the pricing of pharmaceuticals. *Journal of Economics & Management Strategy*, *6*, 75–90.

Gagnon, M.-A. (2013). Corruption of pharmaceutical markets: Addressing the misalignment of financial incentives and public health. *Journal of Law, Medicine & Ethics*, *41*(3), 571–580. https://doi.org/10.1111/jlme.12066

Galizzi, M. M., Ghislandi, S., & Miraldo, M. (2011). Effects of reference pricing in pharmaceutical markets: A review. *PharmacoEconomics*, *29*, 17–33.

Godman, B., Wettermark, B., van Woerkom, M., Fraeyman, J., Alvarez-Madrazo, S., Berg, C., Bishop, I., Bucsics, A., Campbell, S., Finlayson, A. E., Fürst, J., Garuoliene, K., Herholz, H., Kalaba, M., Laius, O., Piessnegger, J., Sermet, C., Schwabe, U., Vlahović-Palčevski, V., . . . Gustafsson, L. L. (2014). Multiple policies to enhance prescribing efficiency for established medicines in Europe with a particular focus on demand-side measures: Findings and future implications. *Frontiers in Pharmacology*, *5*, Article 106. https://doi.org/10.3389/fphar.2014.00106

Grabowski, H. G., & Vernon, J. M. (1992). Brand loyalty, entry, and price competition in pharmaceuticals after the 1984 Drug Act. *The Journal of Law & Economics*, *35*, 331–350.

Gray, A., & Suleman, F. (2015). Pharmaceutical pricing in South Africa. In Z.-U.-D. Babar (Ed.), *Pharmaceutical prices in the 21st century* (pp. 251–265). Adis. https://doi.org/10.1007/978-3-319-12169-7_14

Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, *141*, 61–94. https://doi.org/10.1007/s11205-017-1832-9

Herr, A., & Suppliet, M. (2017). Tiered co-payments, pricing, and demand in reference price markets for pharmaceuticals. *Journal of Health Economics*, *56*, 19–29. https://doi.org/10.1016/j.jhealeco.2017.08.008

Hsiao, F.-Y., Tsai, Y.-W., & Huang, W.-F. (2010). Price regulation, new entry, and information shock on pharmaceutical market in Taiwan: A nationwide data-based study from 2001 to 2004. *BMC Health Services Research*, *10*, Article 218.

Jain, A., Nundy, S., & Abbasi, K. (2014). Corruption: Medicine's dirty open secret. *BMJ*, *348*, Article g4184.

Kaplan, W., & Laing, R. (2005). *Local production of pharmaceuticals: Industrial policy and access to medicines*. World Bank.

Kong, Y., & Seldon, J. R. (2004). Pseudo-generic products and barriers to entry in pharmaceutical markets. *Review of Industrial Organization*, *25*, 71–86.

Levine, D. M., Stephan, D. F., Krehbiel, T. C., & Berenson, M. L. (2008). *Statistics for managers using Microsoft Excel* (5th ed.). Pearson Prentice Hall.

Linden, M. (2013). Generic substitution effects on firm level market shares in the Finnish beta blocker market 1997Q1–2007Q4. *Modern Economy*, *4*, 852–863. https://doi.org/10.4236/me.2013.412092

Manchanda, P., Wittink, D. R., Ching, A., Cleanthous, P., Ding, M., Dong, X. J., Leeflang, P., Misra, S., Mizik, N., Narayanan, S., Steenburgh, T., Wieringa, J., Wosinska, M., & Xie, Y. (2005). Understanding firm, physician and consumer choice behavior in the pharmaceutical industry. *Marketing Letters*, *16*(3), 293–308. https://doi.org/10.1007/s11002-005-5893-1

Marketing Code Authority of South Africa. (2014). *SA code of marketing practice: For the marketing and promotion of medicines, medical devices and in vitro diagnostics*.

Marschak, J. (1950). Rational behavior, uncertain prospects, and measurable utility. *Econometrica*, *18*, 111–141.

Matsumoto, A., Merlone, U., & Szidarovszky, F. (2012). Some notes on applying the Herfindahl–Hirschman index. *Applied Economics Letters*, *19*, 181–184.

Morgan, S. G., Vogler, S., & Wagner, A. K. (2017). Payers' experiences with confidential pharmaceutical price discounts: A survey of public and statutory health systems in North America, Europe, and Australasia. *Health Policy*, *121*, 354–362.

Motta, M. (2004). *Competition policy: Theory and practice*. Cambridge University Press.

National Department of Health. (1996). *National drug policy*. https://www.gov.za/documents/other/national-drugs-policy-01-jan-1996

Nguyen, T. A., Knight, R., Roughead, E. E., Brooks, G., & Mant, A. (2015). Policy options for pharmaceutical pricing and purchasing: Issues for low- and middle-income countries. *Health Policy and Planning*, *30*(2), 267–280. https://doi.org/10.1093/heapol/czt105

Organisation for Economic Co-operation and Development. (2008). *Handbook on constructing composite indicators: Methodology and user guide*.

Organisation for Economic Co-operation and Development. (2018). *Pharmaceutical innovation and access to medicines*.

Organisation for Economic Co-operation and Development. (2020). *Addressing challenges in access to oncology medicines: Analytical report*.

Perišić, A. (2015). Data-driven weights and restrictions in the construction of composite indicators. *Croatian Operational Research Review*, *6*(1), 29–42.

Perumal-Pillay, V. A., & Suleman, F. (2020). Understanding the decision-making process of selection of medicines in the private sector in South Africa. *Journal of Pharmaceutical Policy and Practice*, *13*, Article 17. https://doi.org/10.1186/s40545-020-00223-5

Republic of South Africa. (2024). *National Health Insurance Act, No. 20 of 2023* (Government Gazette No. 50664). https://lawlibrary.org.za/akn/za/act/2023/20/eng@2024-05-16

Rodrigues, V., Gonçalves, R., & Vasconcelos, H. (2014). Anti-competitive impact of pseudo-generics. *Journal of Industry, Competition and Trade*, *14*, 83–98.

Roy, N., Madhiwalla, N., & Pai, S. A. (2007). Drug promotional practices in Mumbai: A qualitative study. *Indian Journal of Medical Ethics*, *4*, 57–61.

Santeramo, F. G. (2017). Methodological challenges in building composite indexes: Linking theory to practice. In V. Jeremic, Z. Radojicic, & M. Dobrota (Eds.), *Emerging trends in the development and application of composite indicators* (pp. 127–139). IGI Global. https://doi.org/10.4018/978-1-5225-0714-7.ch006

Shrestha, S., Poudel, R. S., KC, B., Poudel, B. K., Sapkota, B., Sharma, S., & Khadka, A. (2020). Price variation among different brands of anticancer medicines available in hospital pharmacies of Nepal. *Journal of Pharmaceutical Policy and Practice*, *13*, Article 6. https://doi.org/10.1186/s40545-020-0203-0

Transparency International. (2016). *Corruption in the pharmaceutical sector: Diagnosing the challenges*.

U.S. Department of Justice & Federal Trade Commission. (2010). *Horizontal merger guidelines*. https://www.justice.gov/atr/horizontal-merger-guidelines-08192010

van den Heever, A. (2012). *Review of competition in the South African health system* [Report produced for the Competition Commission].

Van der Merwe Smit, E., & Bredenkamp, J. (2013). Originator and generic medicine: Pricing and market share. *International Journal of Pharmaceutical and Healthcare Marketing*, *7*, 104–119.

Vogler, S., Zimmermann, N., & Babar, Z.-U.-D. (2017). Price comparison of high-cost originator medicines in European countries. *Expert Review of Pharmacoeconomics & Outcomes Research*, *17*, 221–230.

Vogler, S., Zimmermann, N., Leopold, C., Habl, C., & Mazag, J. (2013). Discounts and rebates granted for medicines for hospital use in five European countries. *The Open Pharmacoeconomics & Health Economics Journal*, *5*, 1–10. https://doi.org/10.2174/1876824501305010001

Williams, K. (2007). Pharmaceutical price regulation. *South African Journal on Human Rights*, *23*(1), 1–33. https://doi.org/10.1080/19962126.2007.11864909

World Health Organization. (2020). *WHO guideline on country pharmaceutical pricing policies*.

Wouters, O. J., Kanavos, P. G., & McKee, M. (2017). Comparing generic drug markets in Europe and the United States: Prices, volumes, and spending. *The Milbank Quarterly*, *95*, 554–601.

Zeng, W. (2013). A price and use comparison of generic versus originator cardiovascular medicines: A hospital study in Chongqing, China. *BMC Health Services Research*, *13*, Article 390.

DECLARATIONS

Acknowledgements

The authors would like to thank the Government Employees Medical Scheme (GEMS), the National Department of Health, the Competition Commission of South Africa, and the University of the Witwatersrand for their support and contribution to this work. The views expressed in this manuscript are those of the authors and do not necessarily reflect the official positions of their affiliated institutions.

Author contributions

Molaudi Gaula: Conceptualisation; Methodology; Investigation; Formal analysis; Data interpretation; Writing – original draft; Writing – review and editing; Project administration.

Alex van den Heever: Conceptualisation; Methodology; Review of economic theory, indicator development, interpretation, and industry perspectives; Writing – review and editing.

Ntobeko Mpanza: Regulatory review; Policy interpretation; Writing – review and editing.

Neo Kitsa: Regulatory review; Pricing policy review; Writing – review and editing.

Yongama Njisane: Literature review; Economic interpretation of indicators; Competition and market analysis perspectives; Writing – review and editing.

Selaelo Mametja: Managed care and medical scheme perspectives; Benefit design interpretation; Writing – review and editing.

Mthokozisi Khoza: Data curation; Data extraction; Data mapping; Indicator construction; Preliminary analysis; Writing – review and editing.

Mncengeli Sibanda: Technical review; Regulatory and medical scheme perspectives; Writing – review and editing.

Motale Madigoe: Managed care and medical scheme perspectives; Benefit design and formulary management interpretation; Assessment of implications for tiered formulary design; Writing – review and editing.

Stanley Moloabi: Managed care and medical scheme perspectives; Benefit design interpretation; Writing – review and editing.

All authors contributed to the review and approval of the final manuscript and agree to be accountable for all aspects of the work.

Author affiliations

1. Government Employees Medical Scheme (GEMS), Pretoria, South Africa

2. School of Governance, University of the Witwatersrand, Johannesburg, South Africa

3. National Department of Health, Pretoria, South Africa

4. Pricing Committee, National Department of Health, Pretoria, South Africa

5. Competition Commission of South Africa, Pretoria, South Africa

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request, subject to applicable ethical approvals, data-sharing agreements, and institutional governance requirements. Data availability is restricted due to the confidential and proprietary nature of the underlying pharmaceutical sales and claims datasets.

Declaration of competing interests

The authors include employees of the Government Employees Medical Scheme, whose claims data were analysed, and authors with affiliations or roles at the National Department of Health, its Pricing Committee and the Competition Commission of South Africa. These affiliations and roles constitute relevant non-financial interests. The views expressed are those of the authors and do not necessarily represent their institutions' official positions. The authors declare no competing financial interests.

Ethics statement

This study used secondary, de-identified pharmaceutical sales and medical scheme claims data. No individual patients were contacted, and no personally identifiable information was accessed during the study. The research was conducted in accordance with applicable ethical principles, institutional requirements, and accepted standards of scientific integrity and responsible research conduct.

Ethical approval was obtained from the Pharma-Ethics Research Ethics Committee (Pharma-Ethics REC) under Ethics Reference Number 230225374. Annual re-approval for the research programme was granted on 14 October 2025.

Declaration of AI-assisted technologies

The authors used Microsoft Copilot and Grammarly solely to support language refinement, document structuring, grammar checking, and editorial review. No AI tools were used to generate study data, conduct analyses, interpret findings, or make scientific conclusions. All outputs were critically reviewed, revised, and verified by the authors, who accept full responsibility for the accuracy, integrity, and final content of the manuscript.

Funding

This study received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.



STREAM III

VALUE-BASED CONTRACTING AND PRICING REFORM

DIABETES SHARED VALUE MODEL

Jane Ball, Claudia Pastellides, Nikash Bagirathi and Xolani Masombuka

Published 19 March 2025. *NEJM Catalyst Innovations in Care Delivery*, *6*(4). https://doi.org/10.1056/CAT.24.0517. Copyright © 2025. Abstract reproduced from the original publication. **How to cite this article:** Please cite the original publication: Ball, J., Pastellides, C., Bagirathi, N., & Masombuka, X. (2025). Diabetes shared value model. *NEJM Catalyst Innovations in Care Delivery*, *6*(4). https://doi.org/10.1056/CAT.24.0517

ABSTRACT

Background: Type 2 diabetes is an increasingly prevalent condition among Discovery Health Medical Scheme (DHMS) members, rising from 2.8% of the Scheme population in 2014 to 5.2% in 2024. Costs associated with members living with diabetes also grew substantially, driven largely by hospital admissions. By 2020, care quality indicators showed gaps: only 56% of members had an annual HbA1c test, and just 65% achieved in-range results. To address rising costs, fragmented care and poor outcomes, Discovery Health introduced the Diabetes Shared-Value Programme (Diabetes Care Programme) in 2021.

Methods: The analysis evaluated 80,719 DHMS members over a four-year period (January 2021 to December 2024). The Programme invited eligible members to enrol with a nominated primary care provider responsible for coordinated diabetes management. Key metrics assessed included HbA1c testing rates, disease stage stability, hospital admissions and associated costs.

Results: Enrolment increased from 38% of eligible members in 2021 to 47% in 2024, supported by 1,832 participating providers. Enrolled members achieved significantly better outcomes than non-participants. HbA1c testing was 82% among enrolled members versus 59% among non-enrolled members. In-range HbA1c rates increased from 58% (2020) to 64% (2024), with enrolled members outperforming others by up to 3%. Enrolled members experienced 15% fewer admissions and 14% shorter hospital stays in 2024. By 2023, disease stability rates were 6.4% higher among enrolled members, generating an overall 1% cost saving. These savings enabled a 14% increase in practitioner remuneration, with 63% of the savings allocated back to providers.

Conclusions: The Diabetes Shared-Value Programme effectively improved diabetes care, demonstrating that regular monitoring, proactive management and aligned incentives enhance clinical outcomes and reduce costs. Sustained glycaemic control and disease stability were strongly associated with lower hospital utilisation and total healthcare expenditure. The Programme provides a scalable model for similar shared-value interventions.



STREAM IV

DIGITAL HEALTH AND DATA-DRIVEN PREVENTION

DATA FOR VALUE: INTEGRATING CLINICAL AND SOCIAL RISK DATA FOR POPULATION HEALTH MANAGEMENT

Shannon Knight, Brian Ruff, Riedwaan Jabaar and Prince Chikezie

How to cite this article: Knight, S., Ruff, B., Jabaar, R., & Chikezie, P. (2026). Data for value: Integrating clinical and social risk data for population health management. *Council for Medical Schemes Research Journal*, *1*(1), 122–136.

ABSTRACT

Background: Value-based healthcare requires data infrastructure that can support risk stratification, performance measurement and clinical decision-making at patient, practice and programme levels. Most healthcare systems operate with fragmented data sources, risk adjustment tools confined to diagnosis-linked claims, and digital platforms built for administrative rather than clinical use. The gap between what value-based contracts require and what conventional data systems deliver is the practical obstacle to implementation.

Methods: Over a decade, a healthcare services management organisation in South Africa developed an integrated data architecture combining exogenous diagnosis-linked claims data from medical schemes and laboratories with endogenous data captured through direct clinical assessment. Case-mix tools include Diagnosis-Related Groups for hospital admissions and a novel diagnostic-based Episode Population Grouper for population-level risk. These combine with disease severity measures, psychosocial factors and functional status captured through the organisation's patient management system, producing combined clinical and social risk scores that inform contracting, resource allocation, patient care and performance measurement.

Results: Case-mix adjustment using a novel Episode Population Grouper combined with age and gender explains a substantial proportion of regional variation in admission rates, beyond what models using age, gender and chronic illness count alone achieve. In internal programme data, patients with both high clinical and high social or functional risk experienced substantially higher bed-day rates than patients with high clinical risk alone. Combining clinical and social risk factors identifies patient cohorts that diagnosis-only models may underestimate or miss. Clinical homogeneity within risk groupings supports clinical workflow, not payment calculation alone.

Conclusions: Integrated data systems combining diagnosis-linked claims with direct clinical assessment can be implemented in resource-constrained settings when designed for clinical action. The approach offers a practical model for value-based contracting beyond diagnosis-only risk adjustment.

KEYWORDS

Health data; case mix; risk adjustment; population health management; clinical decision support; value-based care

INTRODUCTION

The data challenge in value-based care

Value-based healthcare aligns payment with outcomes, rewarding quality and efficiency over volume. Realising this requires data infrastructure that supports three functions: stratifying populations by risk, measuring performance fairly by adjusting for case mix, and providing clinical decision support at the point of care (Larsson et al., 2022; Porter et al., 2016; Porter & Lee, 2013). However, most healthcare systems rely on data architectures built for billing and administration, not value-based care. Data sits in silos across payers, providers, laboratories, and hospitals. Risk adjustment tools, where they exist, capture only limited clinical complexity and ignore the disability and social factors that research shows drive outcomes (Buntin & Ayanian, 2017; Hood et al., 2016). Digital platforms, even sophisticated electronic health records, burden clinicians with data entry while providing little decision support in return (Fisher, 2020; McClellan et al., 2025).

The risk adjustment gap

Case-mix-based risk adjustment is foundational to value-based care (Hughes et al., 2004; Pope et al., 2004). Without it, comparisons between providers or populations are not meaningful; apparent differences in outcomes may reflect differences in patient complexity rather than care quality. Yet conventional approaches face considerable limitations. Diagnosis-based models such as the CMS Hierarchical Condition Categories (HCC) used in US Medicare explain only about 10–20% of individual cost variation in prospective models, depending on the population (Ash et al., 2000; Pope et al., 2004). Adding social and functional factors can improve predictive accuracy, yet these factors remain largely absent from operational systems (Chang et al., 2021; Hatef et al., 2019; Joynt et al., 2017). Moreover, they identify subcohorts amenable to different interventions, which is of crucial operational importance.

Paper purpose

This paper describes the experience of a healthcare services management organisation in South Africa that has spent a decade developing an integrated data architecture for value-based primary care. To our knowledge, this is the first operational implementation of integrated diagnosis plus clinical and social risk data architecture at scale in South Africa, built without the decades of public-sector investment in case-mix methodology that have shaped international value-based care efforts elsewhere, including in Europe and North America (Larsson et al., 2022; Porter et al., 2016; Porter & Lee, 2013). We present a practice account: lessons learned through implementation about what data is needed, how it should be organised, and how it can drive clinical action rather than just administrative reporting. The hub-based care model through which this data architecture is deployed, and its programme-level outcomes, are described in a companion article in this issue (Nkwagatse & Richards, 2026).

DATA ARCHITECTURE

Three levels of data

Our data architecture comprises a three-layer data pipeline (see Figure 1). The data pipeline moves from raw ingestion (Bronze) through grouped clinical intelligence (Silver) to strategic value outputs (Gold). These outputs serve three audiences with distinct information needs: patient, practice, and programme. At the patient level, granular data supports clinical decision-making by the care team. At the practice or hub level, aggregated data guides local management: which patients need attention, how the team is performing, where resources should go. At the programme level, data supports strategic management, contract performance and continuous improvement. The levels differ in more than aggregation: update cadences

range from point-of-care currency at the patient level, to daily or weekly at the practice level, to monthly or quarterly at the programme level.

Exogenous data sources

We distinguish exogenous data (received from external sources) from endogenous data (captured within our programme through direct patient interaction). Exogenous sources provide the foundation of our architecture: they show what happens to our patients across the wider healthcare system, beyond the services we deliver. Medical scheme claims data is the primary exogenous source, covering professional claims across specialties, pathology and radiology, medication dispensing, hospital admissions, and surgical and investigative procedures. This data extends back to 2006 for our primary funder population, giving longitudinal visibility over nearly two decades. Claims support attribution (matching patients to practices), utilisation tracking, and outcome measurement through the ICD-10 diagnosis and procedure codes attached to each claim line. Hospital pre-authorisation data, received daily, provides early warning of admissions and enables proactive care coordination. Laboratory results, received through the scheme or directly from pathology providers, allow tracking of disease-control measures such as HbA1c for diabetes, lipid panels for cardiovascular risk, and renal function for kidney disease.

Endogenous data sources

Exogenous data, while valuable, has fundamental limitations. Claims data tells us what services were rendered, not how the patient is doing. Diagnosis codes indicate conditions, not their current severity or trajectory. Lab results measure biological parameters, not functional capacity or social circumstances. To address these limitations, we systematically capture endogenous data through clinical assessment. Our Whole Patient Assessment captures information that claims data cannot provide, including condition severity, functional status using validated instruments, as well as frailty, depression screening, cognitive assessment, social isolation, and economic circumstances. Care coordinators and allied professionals update this record through ongoing patient interaction, so it evolves with the patient. Endogenous data serves a different purpose. Where claims data reveals what happened (the patient was admitted, the test was ordered), endogenous data reveals why and what it means. A patient with repeat admissions for heart failure may have poorly controlled disease or may have good disease control but inadequate social support driving decompensation. Endogenous data can distinguish these two situations and their different interventions. With this two-stream foundation in place, the next requirement is a case-mix framework that can use both streams to make meaningful comparisons across patients, practices and regions.

CASE MIX AND RISK ADJUSTMENT

The central principle

Risk adjustment is not an administrative nicety but a clinical necessity. Without it, value-based care cannot function. A practice serving complex elderly patients will always show higher costs and worse crude outcomes than one serving healthy young adults. Comparing them without adjustment is not meaningful; rewarding or penalising them based on unadjusted metrics is unfair. Our approach rests on a central principle: value (V) is the quality of outcomes (Q) relative to their cost (C), V = Q/C (Figure 1), operationalised as Actual divided by Expected. For any metric (admission rate, bed days, cost) we calculate expected performance given patient complexity and compare actual performance against it. The ratio reveals whether the practice is performing better or worse than predicted; for these metrics, a ratio below one is favourable. This risk-adjusted comparison, not crude rates, drives performance assessment and value-based payment.

Diagnosis-Related Groups (DRGs) to characterise hospital admissions

For hospital admissions, we use Diagnosis-Related Groups (DRGs) as the standard case-mix tool (Annear et al., 2018; Busse et al., 2013; Fetter et al., 1980). DRGs classify inpatient episodes into clinically coherent categories based on the principal diagnosis determined after discharge, procedures, comorbidities, age and sex, with each category carrying a relative weight that reflects expected resource use (Annear et al., 2018; Busse et al., 2013; Fetter et al., 1980). This allows admission rates, lengths of stay and costs to be compared across practices and populations on a like-for-like basis, a prerequisite for any meaningful performance assessment in a value-based contract (Annear et al., 2018; Busse et al., 2013; Fetter et al., 1980). But DRGs characterise only the hospitalisation event; population-level risk adjustment across ambulatory, chronic and preventive care requires a complementary tool.

The Episode Population Grouper

Defining clinical episodes and tracking patients over time across settings have been identified as persistent gaps in conventional healthcare data infrastructure (Fisher, 2020). Population risk adjustment beyond hospitalisation uses the Episode Population Grouper (EPG) we developed jointly with Insight Actuaries, an actuarial services partner: a diagnostic-driven tool that organises claims data into parallel clinical episodes of care when a patient has multiple morbidities. The EPG processes individual claim lines (including doctor visits, laboratory tests and medication dispensed) and groups related services into coherent episodes, drawing on the population-grouper tradition in the risk-adjustment literature (Hughes et al., 2004).

Approximately 1,600 distinct episode types are defined, each triggered by specific ICD-10 codes and linked to expected service patterns. Episodes vary in duration, from brief acute events to continuous chronic conditions, and they capture disease progression over time. A patient's overall severity increases as episodes accumulate from risk factors to early disease, to established chronic illness, to acute-on-chronic episodes, to organ failure (see Figure 2). This allows the EPG to distinguish patients at different points in their disease journey even when they share the same primary diagnosis. The EPG rolls episode-level information into patient-level risk scores through clinical grouping and statistical modelling.

Episodes cluster into approximately 180 clinical groups, which aggregate into over 20 Body System Chapters and 7 syndromes such as Metabolic Syndrome and Frailty. The output is a Disease Burden Index (DBI), a composite score where values below one indicate lower-than-average expected utilisation and values above one indicate higher (Internal programme data, 2024). Across South Africa's 136 primary regions, models combining the DBI with age and gender explained 69% (ordinary least squares, OLS), 80% (random forest) and 85% (beta regression) of the variation in admission rates, against 59% for age, gender and chronic illness count alone (OLS) (Ruff & Knight, 2025). Out of sample, in one hub (Pretoria North East; just under 8,000 members; risk data from September 2023; admissions October 2023 to September 2024), the regionally trained models predicted admission rates of 16% (OLS) to 28% (beta regression) against an observed 20%, with random forest closest at 23% (Internal programme data, 2024; Ruff & Knight, 2025).

The combined DBI plus age-and-gender model performed better than models using age, gender and chronic illness count alone, but the spread across model specifications underscores the limits of diagnosis-based prediction and motivates the integrated clinical-and-social approach (Internal programme data, 2024). Patient-level prediction is inherently harder and remains a focus of continued model development. The EPG belongs to the tradition of diagnosis-based population groupers such as Clinical Risk Groups and the Johns Hopkins Adjusted Clinical Groups (Chang & Weiner, 2010; Hughes et al., 2004), with the distinction that DBI groupings are clinically homogeneous rather than cost-homogeneous. The technical methodology underpinning the EPG, including regional benchmarking design and the role of Insight Actuaries, is described in detail in an earlier peer-reviewed publication (Ruff & Knight, 2025).

Clinical homogeneity

A critical design principle distinguishes our EPG from alternative approaches: clinical homogeneity within groupings. Each episode type and clinical group contains patients with genuinely similar clinical characteristics and care needs, not merely similar costs. Clinically homogeneous groups can inform daily operational clinical workflow; cost-homogeneous groups cannot. Consider diabetes: a cost-based grouper might place all diabetic patients of similar expense together regardless of disease stage, but a patient newly diagnosed with Type 2 diabetes on oral medication has entirely different care needs and likely comorbidities than a patient with insulin-dependent diabetes, recurrent infections, and early renal impairment, even if their current costs are similar. Our EPG distinguishes these patients, enabling care plans that address their different needs appropriately. This design choice has operational consequences. Clinical homogeneity enables the grouper to accurately trigger care plans, guide care coordinator prioritisation, and inform multidisciplinary team discussions. The grouper becomes a clinical tool, not merely a payment mechanism: clinicians reviewing patient risk profiles find meaningful clinical information, not just administrative categories.

INTEGRATING CLINICAL AND SOCIAL RISK

The limitation of diagnosis-based models

Despite the EPG's strong performance in explaining regional variation in admission rates, diagnosis-based risk adjustment has inherent limitations. Research consistently shows that social and functional factors, including housing stability, food security, transportation access and social support, substantially influence health outcomes, yet diagnosis codes capture none of this (Hood et al., 2016; Magnan, 2017).

Our integrated approach

Our response combines the EPG's diagnostic-based clinical risk with psychosocial and functional assessment captured through our patient management system, detailed below. Using our Whole Patient Assessment and ongoing clinical interaction we capture information that claims data cannot provide. Functional assessment uses validated instruments including the Barthel Index (Shah et al., 1989), Timed Up and Go (TUG; Podsiadlo & Richardson, 1991), grip strength (EWGSOP2; Cruz-Jentoft et al., 2019), and the Mini-Mental State Examination (MMSE; Folstein et al., 1975). Depression screening uses the Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001). Social isolation is assessed using the Lubben Social Network Scale (LSNS; Lubben et al., 2006). Environmental factors capture income source, housing type, living arrangements, and use of assistive devices. Family history, sight and hearing issues, and relevant lifestyle factors complete the picture. This assessment generates a social risk score that complements the EPG's clinical risk score. Patients are characterised on two dimensions: clinical complexity from diagnoses and utilisation, and social/functional vulnerability from direct assessment. Together, these dimensions identify patients at risk for poor outcomes that neither dimension alone would catch.

The combined effect

The value of combining clinical and social risk is illustrated by stratifying patients across the two dimensions (see Figure 3) and examining their utilisation. Patients with high clinical risk but low social/functional risk show moderately elevated bed days, as expected. Patients with low clinical risk but high social/functional risk show similar moderate elevation and would typically be missed in diagnosis-based stratification. But patients with both high clinical and high social/functional risk show considerably elevated bed days. Targeting this combined high-risk cohort improves the return on care coordination effort. Analysis of our enrolled population found that patients with high clinical risk combined with high social/functional vulnerability experienced bed-day rates substantially higher than those of patients with high clinical risk

alone, a pattern observed over multiple years (Internal programme data, 2024). This compounding of clinical and social risk identifies an extreme-need cohort whose utilisation diagnosis-based models may underestimate. These patients require intensive intervention that standard risk stratification would not trigger (Internal programme data, 2024). Predictive modelling using machine learning on a subsample of enrolled patients with additional clinical and psychosocial assessment data identified new risk subcategories within the major segments, with different admission rates (Ruff & Knight, 2025). Feature importance analysis showed that BMI, social isolation (LSNS), depression (PHQ-9), living arrangements, grip strength, TUG, smoking status, Barthel score, assistive device use, and MMSE all contributed beyond what diagnosis-based risk captured. The categories corresponded to meaningful clinical distinctions and extended the discrimination achievable with diagnosis data alone (Internal programme data, 2024). Translating this stratification into routine clinical action requires digital infrastructure that connects risk identification to multidisciplinary workflows; the next section describes how our platforms perform that translation.

DIGITAL PLATFORMS

The patient management system

Our patient management system operationally connects our data architecture to clinical action. Purpose-built for value-based care rather than adapted from billing or administrative systems, the system integrates the data sources described above and translates them into workflow that supports care delivery. The system performs four primary functions. First, administrative functions confirm member enrolment, manage consent, and track programme participation. Second, the system provides a longitudinal patient record synthesised from claims data (using the EPG), together with assessment results and care team notes, accessible at the point of care. Third, the system captures new patient data, including assessment results, visit records and clinical observations, feeding the endogenous data stream. Fourth, and most critically, the workflow function stratifies care coordinator work by patient need, prioritising patients for attention based on combined risk segment, with urgency boosts for selected triggers such as recent admission, new chronic diagnosis, or abnormal pathology result. Customised care plans are triggered automatically.

Care plans generated by the system translate risk stratification into concrete clinical action. When a patient's combined risk profile indicates high complexity, the system generates appropriate tasks: diagnostic workup, investigation orders, and management steps. Care coordinators work from dynamically generated task lists that ensure the right patients receive attention at the right time. The system also surfaces cases for multidisciplinary team review.

Analytics platform

While the system supports frontline clinical operations, our analytics platform supports programme-level management. It integrates data from all sources, the endogenous stream from our patient management system plus exogenous data from schemes and laboratories, into structured tables. Case-mix groupers and other analytical tools generate the risk scores, performance metrics, and outcome reports that drive programme management. Analyses are examined both descriptively (the full picture of work done by a practice or hub population) and through the 'contribution to the difference' (CTD), a measure of relative efficiency derived from actual versus expected throughput and outcomes.

Emerging capabilities

Our digital infrastructure continues to evolve. Current development priorities include medication audit tools that flag scripts needing clinical review (for example, medications unsuited to a patient's current BMI or renal function). Summary reports generated automatically from patient data can prepare care coordinators for patient contacts by surfacing relevant history and outstanding care gaps. Artificial intelligence (AI) in our

system is being explored in task-specific modes: machine-learning-driven feature identification in combined clinical-and-social risk modelling, large-language-model-supported documentation and summary generation, and medication audit. We do not treat AI as a general-purpose intelligence layer but rather as an aid to research and development where new capabilities are validated task by task against clinical utility before broader deployment.

These AI capabilities extend rather than replace human judgement. The goal is not autonomous clinical decision-making but augmented intelligence that surfaces relevant information, flags concerns, and reduces administrative burden so more time can be spent with patients. Clinicians retain authority over every clinical decision; the tools serve their judgement rather than substitute for it. The challenge is ensuring these tools are genuinely useful and do not produce fabricated or dangerous recommendations.

OPERATIONAL EXPERIENCE

Effective elements

Several elements of our data approach have proved particularly valuable. First, the commitment to purpose-built rather than adapted systems. Starting with clinical requirements and building technology to meet them produced tools that clinicians use, because they address real workflow needs and avoid the legacy issues that characterise EMRs built on administrative systems. Second, the integration of exogenous and endogenous data into unified patient records. Clinicians can see both what services a patient received elsewhere and what assessments revealed during programme interaction, visibility neither source alone could offer.

The integration required substantial technical work but produces operational value daily. Third, the discipline of clinical homogeneity in case-mix tools. Because our EPG groups patients by clinical rather than cost similarity, it supports clinical workflow in ways that cost-based groupers cannot. Care coordinators trust the risk stratification because it corresponds to the clinical reality they observe. Fourth, the explicit capture of social and functional risk factors. Making these factors visible in risk assessment changed clinical practice: care coordinators now routinely assess and address them, knowing they influence outcomes as much as clinical complexity.

Implementation challenges

We learned equally from failures. Early versions of our platforms prioritised comprehensive data capture over usability, producing systems that clinicians found burdensome, and system use diminished. Simplifying interfaces while maintaining essential functionality improved adoption substantially. Additionally, rapid, unanticipated changes in member status made the administration system inaccurate and redundant, and it had to be rebuilt. We also learned that good downstream architecture cannot compensate for poor upstream data. Incomplete pre-authorisation records, missing chronic condition flags, and inconsistent coding from source systems continued to distort outputs regardless of how well we integrated them. Building team awareness of these gaps, knowing when to trust a risk score and when to treat it sceptically, became as important as the data infrastructure itself. Lastly, we underestimated the difficulty of change management. Good data does not automatically produce good decisions. Clinicians accustomed to working without data support needed time and coaching to incorporate it into their practice. The technology was necessary but not sufficient: human factors determined whether data translated into action.

Lessons learned

First, build for clinical action, not reporting. The test of any data element is whether a named person can decide or take an action because of it. Data that fails this test is not useful regardless of its sophistication.

Second, invest early in integration infrastructure. Healthcare data is inherently fragmented across payers, providers and laboratories; the technical and governance work required to unify it into coherent patient records is substantial and consistently underestimated. Third, capture social and functional risk from the start. Retrofitting these measures into a diagnosis-only system is harder than building them in. The extreme-risk cohort whose needs are underestimated by diagnosis-only models is not fully identifiable without them. Fourth, prioritise clinical homogeneity in case-mix groupers. Cost-similarity groupings serve payment but cannot inform clinical workflow; clinical-similarity groupings do both.

DISCUSSION

Policy relevance

For South Africa specifically, this experience offers lessons as the health system evolves towards universal health coverage (UHC). The National Health Insurance Act 20 of 2023 establishes a single-payer NHI Fund as the central purchaser of health services for the country, with strategic purchasing for outcomes a stated policy intent (Republic of South Africa, 2024; Solanki et al., 2025). While our experience is in the private medical scheme environment, the underlying data architecture is sector-agnostic: ICD-10-coded encounter data, structured clinical assessment, and case-mix grouping operate identically whether the funder is a single public payer or multiple private schemes (Fisher, 2020). The implementation challenge is twofold: substantial technical infrastructure investment in interoperable claims systems, ICD-10 coding and DRG discipline, and structured assessment capture; and the organisational discipline to design data flows around clinical decisions rather than billing requirements (Fisher, 2020). More broadly, the experience speaks to middle-income countries where data infrastructure is limited but digital transformation is underway. Starting with clear clinical purposes, building integrated data flows, and incorporating social and functional risk factors offers a path towards data systems that serve value rather than administration.

Limitations

Several limitations constrain generalisation from our experience. We operate with a single medical scheme whose data sharing arrangements may not replicate elsewhere. Our enrolled population differs from other populations, and our decade of development reflects substantial ongoing investment that others may not be able to replicate. Our evidence of impact, while consistent with programme theory, has limitations noted in the risk adjustment literature (Chang et al., 2021; Chang & Weiner, 2010). We cannot conduct randomised trials or exclude outliers because all patients in the programme must be managed to the best of our ability. Selection effects influence comparisons, which is why local measures reflecting the local environment, including unmeasured peculiarities, must form part of practice scoring. We believe our approach creates value; rigorous quantification of that value is an active area of work, with ongoing analyses of risk-adjusted outcomes against contemporary benchmarks. Consistent with the practice-account format, the quantitative findings reported here are descriptive observations from routine programme analyses; formal evaluation with prespecified cohorts, exposure-based rates and measures of uncertainty is the natural next step. Our approach sits between commercial diagnosis-only systems that omit social risk and academic social determinants of health frameworks that lack operational scaffolding for clinical action; this practical positioning is a strength for implementation but a limitation for direct comparability to either tradition (Buntin & Ayanian, 2017; Chang et al., 2021).

Future directions

Our data architecture continues to evolve towards the Learning Health System vision of care in which every clinical encounter both draws on a patient's complete history and contributes to improved future care (Fisher, 2020). Enhanced AI capabilities could accelerate tool development and surface clinical insight at the point of care. Patient-facing tools could extend the system's reach, engaging patients as active participants in their

data and care. Additional data sources such as wearables and home monitoring equipment could enrich the clinical picture, provided raw streams are filtered for clinically meaningful signal and integrated with the existing patient record rather than processed in isolation. Yet we remain mindful that data is a means, not an end. The goal is better patient outcomes at sustainable cost, not more sophisticated analytics for their own sake. Each enhancement must demonstrate contribution to that goal or be questioned. Data for value requires keeping value, not data, as the measure of success.

CONCLUSION

A decade of developing and operating integrated data infrastructure for value-based primary care yields several principal lessons. Data serves value only when it drives clinical action; sophisticated systems that do not change practice create no benefit. Combining clinical and social risk factors identifies cohorts that diagnosis-based models underestimate or miss, particularly extreme-need patients whose risk conventional adjustment understates. Data architecture must operate at three levels (patient, practice, and programme), each requiring distinct tools, metrics, and cadences. Clinical homogeneity in case-mix groupings is essential for tools to support clinical workflow rather than payment calculation alone. Data in value-based care is not fundamentally about technology. It is about information flows that enable better decisions. Technology serves those flows, but the design choices that matter are clinical and organisational, not technical. What information do clinicians need to care for patients effectively? How should populations be stratified to match resources to needs? What factors truly predict outcomes? What feedback enables continuous improvement? When these questions drive design, technology becomes genuinely enabling. When technology leads and clinical questions follow, the result is often sophisticated systems that nobody uses and investments that create no value.

The argument that inadequate information systems are a root cause of persistent performance problems has been made at policy level for international health systems (Fisher, 2020); this account offers one operational realisation of that principle. Healthcare generates enormous volumes of data. Most of it sits unused: captured for compliance, stored without purpose, never translated into insight or action. The opportunity for integrated health systems is not more data but better use of data. This means patient-level data that guide individual care decisions, practice-level data that enable team improvement, and programme-level data that support strategic management. Running through all levels is a recognition that the numbers represent people, their illnesses, their vulnerabilities, their hopes for health. When data serves those people, value-based care fulfils its promise. That is the goal we pursue.

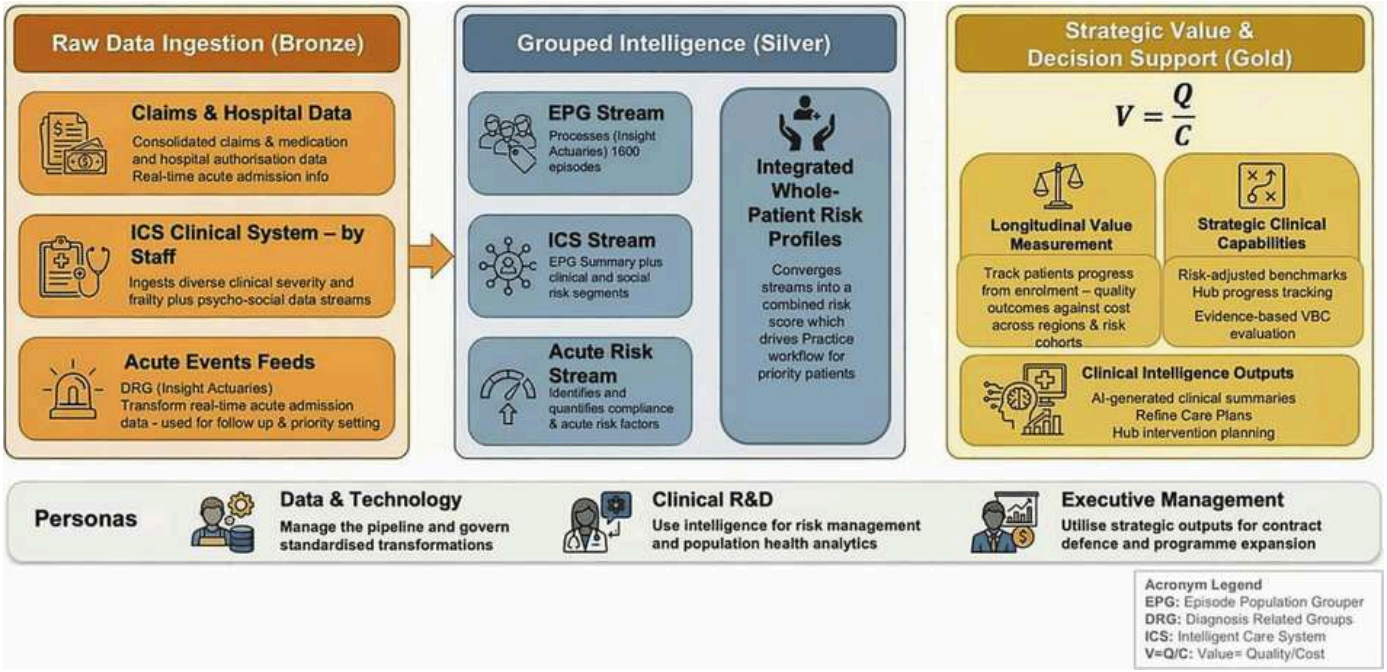

Figure 1: Clinical intelligence architecture

Three-layer data pipeline from raw data ingestion (Bronze) through grouped intelligence (Silver) to strategic value and decision support (Gold). Exogenous claims (including pathology and laboratory results), hospital authorisation and acute event data are combined with endogenous clinical assessment data from the Intelligent Care System (ICS), the programme's patient management and clinical workflow system, to inform patient-, practice- and programme-level decisions. In the Gold layer, value (V = Q/C) is the quality of outcomes (Q) relative to their cost (C), measured as actual relative to expected performance. Personas are the internal teams that operate and use the pipeline. VBC, value-based care.

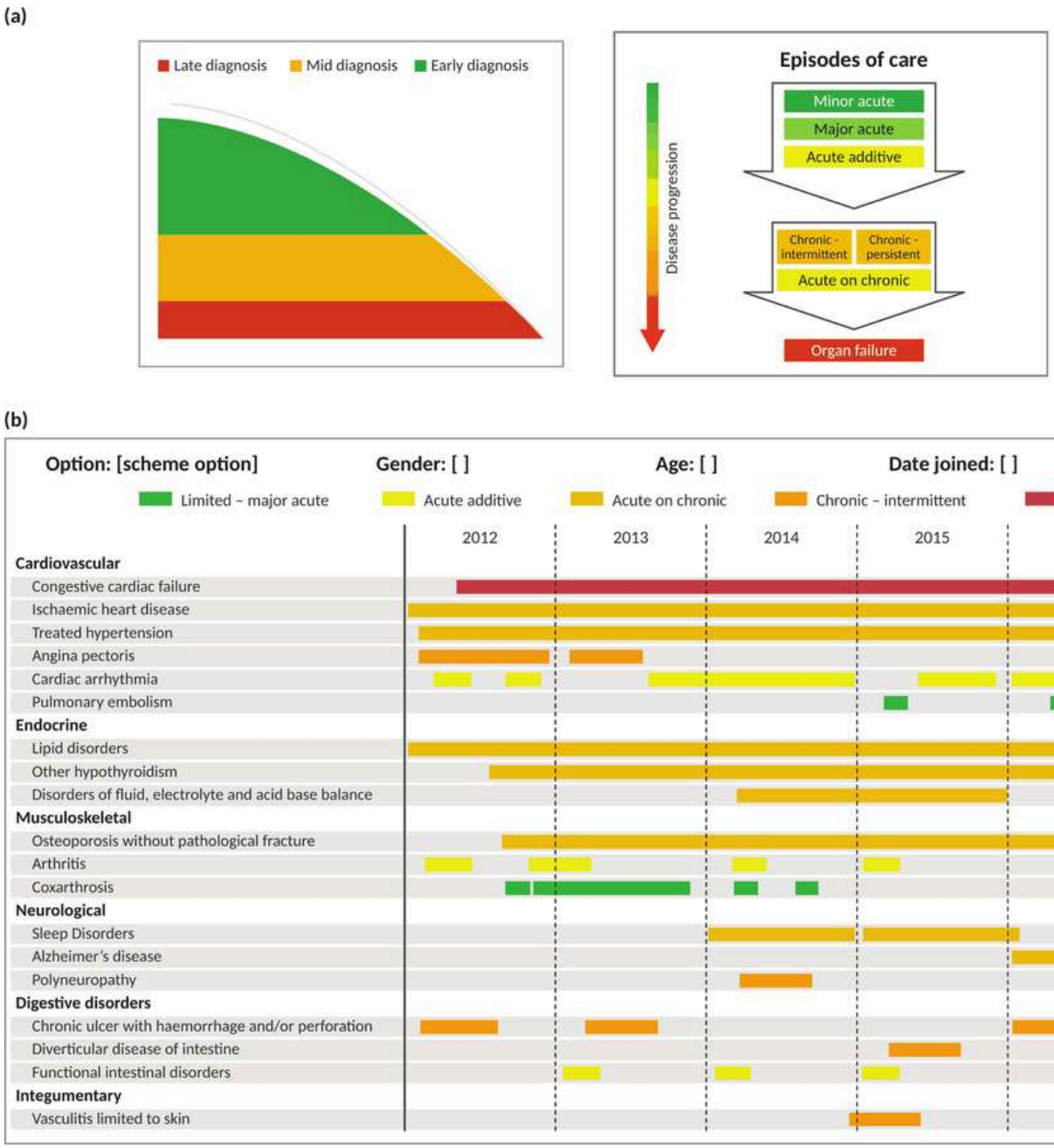

Figure 2: Episode Population Grouper outputs: (a) diagnosis stage and episode-of-care categories; (b) illustrative member episode timeline

Individual claim-line data is organised into approximately 1,600 distinct episode types triggered by ICD-10 codes, clustered into approximately 180 clinical groups, which aggregate into over 20 Body System Chapters and 7 syndromes such as Metabolic Syndrome and Frailty. Episodes range from minor acute events to continuous chronic conditions and capture disease progression over time. The main panel shows an illustrative, synthetic patient timeline (no patient data), with episodes grouped by body system and coloured by episode type; the left-hand panels show how episode types escalate with disease progression, from acute episodes to organ failure.

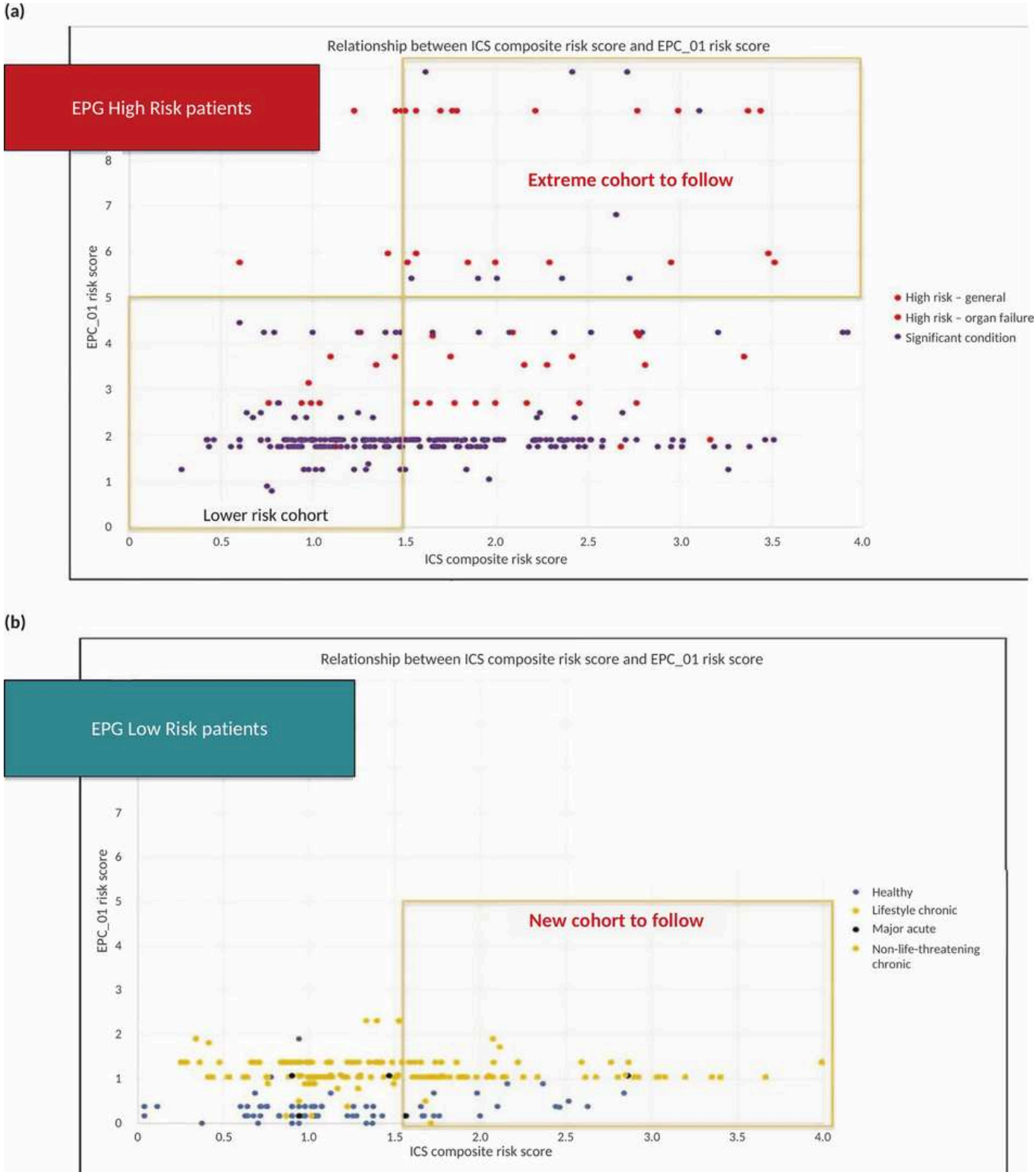

Figure 3: Combining clinical risk with frailty and psychosocial risk for greater risk identification accuracy: (a) EPG high-risk patients; (b) EPG low-risk patients

Two-dimensional stratification using the EPG clinical risk score (Y-axis) and the ICS composite social and functional risk score (X-axis), shown for EPG high-risk patients (upper panel) and EPG low-risk patients (lower panel). The stratification identifies an extreme-need cohort (high clinical and high social/functional risk) and a new cohort with low clinical but high social/functional risk; the utilisation observed in these cohorts is described in the text. Diagnosis-only models may underestimate this compounding of risk.

REFERENCE LIST

Annear, P. L., Kwon, S., Lorenzoni, L., Duckett, S., Huntington, D., Langenbrunner, J. C., Murakami, Y., Shon, C., & Xu, K. (2018). Pathways to DRG-based hospital payment systems in Japan, Korea, and Thailand. *Health Policy*, *122*(7), 707–713. https://doi.org/10.1016/j.healthpol.2018.04.013

Ash, A. S., Ellis, R. P., Pope, G. C., Ayanian, J. Z., Bates, D. W., Burstin, H., Iezzoni, L. I., MacKay, E., & Yu, W. (2000). Using diagnoses to describe populations and predict costs. *Health Care Financing Review*, *21*(3), 7–28.

Buntin, M. B., & Ayanian, J. Z. (2017). Social risk factors and equity in Medicare payment. *New England Journal of Medicine*, *376*(6), 507–510. https://doi.org/10.1056/NEJMp1700081

Busse, R., Geissler, A., Aaviksoo, A., Cots, F., Häkkinen, U., Kobel, C., Mateus, C., Or, Z., O'Reilly, J., Serdén, L., Street, A., Tan, S. S., & Quentin, W. (2013). Diagnosis related groups in Europe: Moving towards transparency, efficiency, and quality in hospitals? *BMJ*, *346*, Article f3197. https://doi.org/10.1136/bmj.f3197

Chang, H.-Y., Hatef, E., Ma, X., Weiner, J. P., & Kharrazi, H. (2021). Impact of Area Deprivation Index on the performance of claims-based risk-adjustment models in predicting health care costs and utilization. *Population Health Management*, *24*(3), 403–411. https://doi.org/10.1089/pop.2020.0135

Chang, H.-Y., & Weiner, J. P. (2010). An in-depth assessment of a diagnosis-based risk adjustment model based on national health insurance claims: The application of the Johns Hopkins Adjusted Clinical Group case-mix system in Taiwan. *BMC Medicine*, *8*, Article 7. https://doi.org/10.1186/1741-7015-8-7

Cruz-Jentoft, A. J., Bahat, G., Bauer, J., Boirie, Y., Bruyère, O., Cederholm, T., Cooper, C., Landi, F., Rolland, Y., Sayer, A. A., Schneider, S. M., Sieber, C. C., Topinková, E., Vandewoude, M., Visser, M., & Zamboni, M. (2019). Sarcopenia: Revised European consensus on definition and diagnosis. *Age and Ageing*, *48*(1), 16–31. https://doi.org/10.1093/ageing/afy169

Fetter, R. B., Shin, Y., Freeman, J. L., Averill, R. F., & Thompson, J. D. (1980). Case mix definition by diagnosis-related groups. *Medical Care*, *18*(2, Suppl.), iii, 1–53.

Fisher, E. S. (2020). Reforming health care: The single system solution. *NEJM Catalyst Innovations in Care Delivery*, *1*(5). https://doi.org/10.1056/CAT.20.0456

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Hatef, E., Weiner, J. P., & Kharrazi, H. (2019). A public health perspective on using electronic health records to address social determinants of health: The potential for a national system of local community health records in the United States. *International Journal of Medical Informatics*, *124*, 86–89. https://doi.org/10.1016/j.ijmedinf.2019.01.012

Hood, C. M., Gennuso, K. P., Swain, G. R., & Catlin, B. B. (2016). County Health Rankings: Relationships between determinant factors and health outcomes. *American Journal of Preventive Medicine*, *50*(2), 129–135. https://doi.org/10.1016/j.amepre.2015.08.024

Hughes, J. S., Averill, R. F., Eisenhandler, J., Goldfield, N. I., Muldoon, J., Neff, J. M., & Gay, J. C. (2004). Clinical Risk Groups (CRGs): A classification system for risk-adjusted capitation-based payment and health care management. *Medical Care*, *42*(1), 81–90. https://doi.org/10.1097/01.mlr.0000102367.93252.70

Internal programme data. (2024). *Claims data analysis, assessment data, and predictive modelling results* [Unpublished internal data]. Professional Provider Organisation Services (Pty) Ltd.

Joynt, K. E., De Lew, N., Sheingold, S. H., Conway, P. H., Goodrich, K., & Epstein, A. M. (2017). Should Medicare value-based purchasing account for social risk? *New England Journal of Medicine*, *376*(6), 510–513. https://doi.org/10.1056/NEJMp1616278

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Larsson, S., Clawson, J., & Kellar, J. (2022). *The patient priority: Solve healthcare's value crisis by measuring and delivering outcomes that matter to patients*. McGraw-Hill.

Lubben, J., Blozik, E., Gillmann, G., Iliffe, S., von Renteln Kruse, W., Beck, J. C., & Stuck, A. E. (2006). Performance of an abbreviated version of the Lubben Social Network Scale among three European community-dwelling older adult populations. *The Gerontologist*, *46*(4), 503–513. https://doi.org/10.1093/geront/46.4.503

Magnan, S. (2017). Social determinants of health 101 for health care: Five plus five. *NAM Perspectives*. https://doi.org/10.31478/201710c

McClellan, M. B., Legner, V., & Debab, S. (2025). Modernizing risk adjustment for value-based care without clinician burnout. *JAMA Health Forum*, *6*(10), Article e254756. https://doi.org/10.1001/jamahealthforum.2025.4756

Nkwagatse, H., & Richards, B. (2026). Implementing integrated primary care hubs in private healthcare: Lessons from a decade of practice in South Africa. *Council for Medical Schemes Research Journal*, *1*(1), 26–38.

Podsiadlo, D., & Richardson, S. (1991). The timed "Up & Go": A test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society*, *39*(2), 142–148. https://doi.org/10.1111/j.1532-5415.1991.tb01616.x

Pope, G. C., Kautter, J., Ellis, R. P., Ash, A. S., Ayanian, J. Z., Iezzoni, L. I., Ingber, M. J., Levy, J. M., & Robst, J. (2004). Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financing Review*, *25*(4), 119–141.

Porter, M. E., Larsson, S., & Lee, T. H. (2016). Standardizing patient outcomes measurement. *New England Journal of Medicine*, *374*(6), 504–506. https://doi.org/10.1056/NEJMp1511701

Porter, M. E., & Lee, T. H. (2013). The strategy that will fix health care. *Harvard Business Review*, *91*(10), 50–70.

Republic of South Africa. (2024). *National Health Insurance Act 20 of 2023* (Government Gazette No. 50664, 16 May 2024). Government Printer. https://www.gov.za/documents/acts/national-health-insurance-act-20-2023-english-afrikaans-16-may-2024

Ruff, B., & Knight, S. (2025). The role of predictive analytics and big data in healthcare decision-making. *Southern African Health Journal*, 6th ed., 30–45. https://www.bhfglobal.com/Publication?report=6617107c-435b-f111-bec6-7ced8d5c37e0

Shah, S., Vanclay, F., & Cooper, B. (1989). Improving the sensitivity of the Barthel Index for stroke rehabilitation. *Journal of Clinical Epidemiology*, *42*(8), 703–709. https://doi.org/10.1016/0895-4356(89)90065-6

Solanki, G., Brijlal, V., Morar, R., Cornell, J., Myburgh, N., & Cleary, S. (2025). The National Health Insurance Act: Possible private health funding reform scenarios. *South African Medical Journal*, *115*(5), Article e2550. https://doi.org/10.7196/SAMJ.2025.v115i5.2550

DECLARATIONS

Acknowledgements

The authors wish to acknowledge the general practitioners, care coordinators, allied health professionals, data engineers, analysts, and management team whose work over more than a decade has made the practice described in this paper possible. We are grateful to the medical scheme partner whose contracting framework enabled the programme. No external research funding was received for the preparation of this manuscript; the underlying programme of work has been funded through value-based contracts between the authors' organisation and a major South African medical scheme.

Author contributions

Shannon Knight: Conceptualisation, Methodology, Writing – Original Draft; Brian Ruff: Conceptualisation, Methodology, Writing – Review & Editing; Riedwaan Jabaar: Conceptualisation, Methodology, Supervision, Writing – Review & Editing; Prince Chikezie: Writing – Review & Editing, Visualisation.

All authors have read and approved the final version of this manuscript.

Data availability statement

The aggregated, de-identified operational data and analytical outputs presented in this paper are derived from internal programme evaluation and are the property of the authors' organisation and its contracting medical scheme partner. Underlying claims, clinical assessment, and case-mix data are governed by contractual data-sharing arrangements that constrain external access. Requests for further data access may be directed to the corresponding author, subject to institutional and contractual data-sharing constraints.

Declaration of competing interest

The authors are employed by Professional Provider Organisation Services (Pty) Ltd, the organisation that designed and delivers the integrated primary care programme described in this paper, under value-based contracts with the medical scheme whose data are referred to. BR and RJ are co-founders of the organisation. The Episode Population Grouper was co-developed by, and is jointly owned by, the authors' organisation and Insight Actuaries, which has licensed it to third parties; the authors' organisation has received modest revenue from these licences. The DRG grouper used by the programme was developed by Insight Actuaries. The ICS platform is the organisation's own system. No other financial conflicts of interest or personal relationships that could have influenced the work are declared.

Declaration of ethics

The authors declare that the writing of this manuscript adheres to ethical standards of publication, in accordance with scientific principles, and is free from plagiarism. The programme described in this paper is delivered by clinicians and an organisation that hold direct ethical responsibility for the care of enrolled patients, operating within the clinical governance and professional standards that bind healthcare practitioners in South Africa. As this article is a practice account of an operational programme rather than a research study, research ethics committee approval was not sought and no waiver was requested; this includes the machine-learning analysis (Ruff & Knight, 2025), which was an operational analysis undertaken to target care within the programme. Patient data are processed by the authors' organisation, as Responsible Party under the Protection of Personal Information Act (POPIA), with the explicit consent of each member obtained at enrolment for purposes that include managing

treatment, monitoring the quality of care and developing healthcare services. Only aggregated, de-identified results are reported, and separate consent to publication was therefore not obtained. The social and functional risk scores are used solely to direct care resources to patients with greater need; the authors' organisation has no role in setting member contributions or benefits, and South African medical schemes are community-rated under the Medical Schemes Act.

Declaration of assistive technologies in the writing process

The authors declare that generative artificial intelligence (GenAI) and AI-assisted tools were used in a limited and responsible manner during the preparation of this manuscript. Specifically, large language models were used for drafting assistance, structural editing, and reference verification. All AI-generated content was critically reviewed, revised, and validated by the authors against original sources to ensure accuracy, originality, and compliance with ethical and scholarly standards. All technical content, analytical interpretations and conclusions are the work of the authors. The authors take full responsibility for the final content of this manuscript.



STREAM IV

DIGITAL HEALTH AND DATA-DRIVEN PREVENTION

IMPROVED SLEEP BEHAVIOURS AND HEALTHCARE COSTS AMONGST MEMBERS OF THE DISCOVERY HEALTH MEDICAL SCHEME: A LONGITUDINAL OBSERVATIONAL STUDY

Katherine Tregoning, Simon Henrie, Deepak Patel, Cassi Rubin, Lara Wayburne, Claudia Pastellides, Roseanne Harris and Mosima Mabunda

How to cite this article: Tregoning, K., Henrie, S., Patel, D., Rubin, C., Wayburne, L., Pastellides, C., Harris, R., & Mabunda, M. (2026). Improved sleep behaviours and healthcare costs amongst members of the Discovery Health Medical Scheme: A longitudinal observational study. *Council for Medical Schemes Research Journal*, *1*(1), 137–149.

ABSTRACT

Background: Sleep is fundamental to health, yet inadequate sleep remains widespread and is associated with increased risk of disease and mortality. Its relationship with healthcare costs is less studied. The objective of this study was to establish the association between changes in sleep behaviour and healthcare costs.

Methods: We conducted a longitudinal observational cohort study of 28,808 South African medical scheme members with wearable-tracked sleep data. A composite Sleep Habit Score (0–100) incorporating sleep duration, regularity and quality was used to classify members into "Good" (≥75), "Medium" (50–74) or "Poor"

(<50) sleep habit bands. Baseline sleep habits and covariates were measured in 2022, sleep habit changes were observed in 2023, and in-hospital claims costs were evaluated in 2024. Adjustments were made for baseline health status and confounders using both a gradient-boosted risk adjustment model and an X-learner causal inference model.

Results: Amongst 22,700 baseline members in the Poor or Medium sleep bands, 3,400 (15%) improved their sleep habit in 2023. Those who improved had an estimated ~7% lower in-hospital claims costs in 2024 than those who did not (risk-adjustment model: 95% CI −14.2% to +0.1%). Members who maintained a Good sleep habit had ~5% lower in-hospital healthcare costs than those whose sleep habit deteriorated (X-learner model: 95% CI −9.9% to −0.5%).

Conclusions: In this population of medical scheme members, improved sleep habits over one year were associated with an estimated ~7% reduction in subsequent in-hospital costs, but both 95% confidence intervals included zero, so the primary hypothesis was not statistically supported at the 5% level. These observational findings support sleep health as a potential lever for reducing healthcare demand.

KEYWORDS

Sleep; healthcare costs; wearable devices; longitudinal observational study; preventive health; causal inference

INTRODUCTION

Sleep is increasingly recognised as a key determinant of health and wellbeing (National Sleep Foundation, 2025). In the United States, about one in three adults does not obtain sufficient sleep (Liu et al., 2016), and modern lifestyles, including artificial light exposure, digital device use and an "always-on" work culture, erode sleep duration and regularity, two fundamental components of sleep architecture (Giuntella & Mazzonna, 2019). Chronic poor sleep has pervasive health consequences, including higher risk of cardiometabolic diseases such as hypertension (Bock et al., 2022; Grandner et al., 2018), diabetes (Antza et al., 2021; Grandner et al., 2016), depression (Irwin, 2015), motor-vehicle accidents (Philip et al., 2014), and premature mortality (Cappuccio et al., 2010; Windred et al., 2024; Yin et al., 2017). Recent US survey data (National Sleep Foundation, 2025) indicate that 60% of adults do not regularly get the recommended 7 to 9 hours of sleep, and almost 50% of adults have trouble staying asleep for 3 or more nights per week. These trends highlight a sleep health crisis that parallels the better-known epidemics of physical inactivity and poor diet.

Sleep deficits can plausibly increase healthcare needs by exacerbating chronic conditions and elevating risk of acute illness events, yet few population studies have directly quantified impacts on healthcare utilisation or costs. Given the global underprioritisation of healthy sleep in healthcare and society, there is a need to evaluate whether improving sleep behaviours might translate into tangible reductions in healthcare costs. Our primary research question is whether improvements in sleep habits are associated with lower in-hospital healthcare expenditure in a large adult medical scheme population. We hypothesised that members who improve their sleep habits (across duration, regularity, and quality) would exhibit lower subsequent in-hospital claims costs compared with those whose sleep habits did not improve, after adjusting for baseline health status and confounders.

By leveraging a large longitudinal dataset, integrated with health insurance claims and demographic data, this study aimed to quantify the association between sleep and healthcare costs. This analysis treats improvements in sleep habits as a preventive intervention and uses observational methods to simulate the effect of better sleep on healthcare costs. Our contribution lies in providing real-world evidence of the potential incremental value of improved sleep, in terms of avoidable in-hospital costs, in a private medical scheme context.

LITERATURE REVIEW

To date, large epidemiological studies affirm that sleep is fundamental to health, and inadequate sleep is associated with a broad range of adverse health outcomes (Watson et al., 2015). The American Heart Association recently added sleep health to the original seven health behaviours and factors that it associates with reduced risk of cardiovascular disease, longevity and higher quality of life. The 8 components, called Life's Essential 8, now include healthy sleep alongside eating better, participation in physical activity (PA), not smoking, healthy weight, in-range fasting blood glucose, controlled blood lipids, and in-range blood pressure (Lloyd-Jones et al., 2022).

An analysis of prospective cohort studies shows approximately 7 hours per day is associated with the lowest risk of mortality and cardiovascular disease. Every hour below 7 hours increases risk by roughly 6% for all-cause mortality and cardiovascular disease (Yin et al., 2017). Short-duration sleep increases the risk of hypertension by about 7% overall, and about 11% if sleep is under 5 hours (Hosseini et al., 2024). Several meta-analyses link short duration or poor sleep to a higher risk of type 2 diabetes (Lu et al., 2021; Wang et al., 2023). One meta-analysis demonstrated that short sleep duration had an increased relative risk of type 2 diabetes (RR = 1.22, 95% CI: 1.15–1.29, $p < 0.001$) compared with normal-duration sleep (Lu et al., 2021).

"The risk of traffic accidents was 3.35 times higher in subjects who reported very poor quality sleep during the last 3 months (CI 95%: 1.30–8.63, p<0.05) and 1.69 times higher in subjects reporting sleeping 6 hours or fewer per night during the last 3 months (CI 95%: 1.00–2.85, p<0.05)" (Philip et al., 2014).

Data on sleep patterns in South African cohorts remains limited. Kim and Madhavan (2024) used data from the South African Time Use Study to examine the effects of gender, employment status and income on sleep. They found that employed individuals were more likely to achieve the recommended amount of sleep. However, employed women were more likely to experience short sleep compared to unemployed women. Men with employed spouses were nearly twice as likely to report short sleep compared to those whose spouses were unemployed. Individuals in the highest income group were more likely to obtain the recommended amount of sleep.

In terms of healthcare utilisation, prior research has focused on patients with diagnosed sleep disorders (Olatunde & Patton, 2025). For example, sleep disorders are associated with significantly higher healthcare costs in the United States, with one analysis estimating an extra $95 billion per year nationwide due to conditions such as insomnia and sleep apnoea (Huyett & Bhattacharyya, 2021). People with untreated insomnia were observed to have higher healthcare utilisation and costs, with the excess highest for inpatient care (rate ratio [RR] 1.61; 95% confidence interval [CI] 1.59–1.64) (Wickwire et al., 2019). Similarly, employees who had poor or short-duration sleep were more likely to be absent from work (all $p < 0.0005$), have lower work performance ratings (all $p < 0.0005$), and have higher healthcare costs ($p < 0.0005$) (Hui & Grandner, 2015).

However, there is a knowledge gap on whether improving sleep behaviour can reduce health service utilisation, since most evidence is either cross-sectional or focuses on clinical sleep disorder management. This study aims to address this gap by examining longitudinal associations between sleep improvements and healthcare claims in a real-world South African cohort.

METHODOLOGY

Study design and population

The study employed a longitudinal observational design to examine whether changes in sleep behaviour were associated with downstream differences in hospital costs among adult members of Discovery Health Medical Scheme (DHMS) who were also enrolled in the Vitality wellness programme and who tracked their sleep with a wearable device. The analytic dataset integrated healthcare claims data, demographic data and Discovery Vitality programme engagement with linked daily sleep-tracking measurements obtained from wearable devices.

Measurement of sleep behaviour

Sleep behaviour was characterised using multiple nightly metrics captured by wearable devices, including nightly duration (total hours slept), sleep timing regularity (consistency of sleep onset time), and sleep quality proxied by the proportion of time in device-defined "deep" and rapid eye movement (REM) sleep stages.

To summarise these dimensions, we constructed a composite Sleep Habit Score ranging from 0 to 100, designed to capture multidimensional sleep behaviour in a single measure. The score was derived using a machine learning approach (gradient boosting XGBoost model) that combined key sleep metrics (duration, regularity, %REM, %deep) with demographic and baseline health characteristics to generate a risk-predictive index of sleep-related health status to predict in-hospital healthcare costs. For each individual, the 'Sleep Habit Score' was calculated as the rolling 10-week average of the daily score, to reduce the effect of night-

to-night fluctuation and reflect sustained sleep behaviour. Based on the composite Sleep Habit Score, members were categorised into three sleep habit bands as shown below in Table 1.

Table 1: Sleep habit bands, Sleep Habit Score and population distribution

Sleep habit band	Average rolling 10-week score	Number of members in 2023
Good	>75	10,964 (18%)
Medium	50–75	37,106 (61%)
Poor	≤50	12,743 (21%)

Thresholds were defined to represent meaningful differences in sleep health. For example, a 'Good' habit corresponds to averaging roughly 7–8 hours of sleep with regular sleep start times and high sleep quality (defined by time spent in deep and REM sleep), whereas a 'Poor' sleep habit indicates persistently short, irregular, or low-quality sleep.

Time periods

The research followed a three-period design from 2022 to 2024, with the following time horizons:

- Baseline period (2022): measurement of demographic characteristics, baseline healthcare utilisation and initial sleep behaviour.
- Exposure assessment period (2023): classification of changes in sleep behaviour, including improvement, maintenance or deterioration in sleep habit band.
- Outcome period (2024): assessment of in-hospital healthcare claims costs.

This structure allows for evaluation of the association between changes in sleep behaviour and subsequent healthcare costs while accounting for baseline differences.

To ensure data completeness and minimise bias arising from membership discontinuity and sparse sleep tracking, the study population was restricted to members with continuous medical scheme enrolment from 2022 through 2024. In addition, members were required to have valid wearable-derived sleep habit data in both 2022 and 2023, defined as more than 60 tracked days within a 10-week period in each year. Sleep data were limited to observations recorded using an Apple or Garmin wearable device to maintain measurement consistency across participants. After applying these criteria, 28,808 members were included in the final analytic cohort.

Outcome measures

The primary outcome measure was in-hospital healthcare costs in 2024, defined as total annual in-hospital claims cost per member in South African rand (ZAR). These costs included expenses associated with hospital admissions and related in-hospital services incurred during the admission period. Planned (elective) procedures and major trauma admissions were excluded to reduce confounding events less directly associated with behavioural or preventative health factors.

Analytical approach and population considerations

The relationship between sleep behaviour and health outcomes is complex due to bidirectional effects. Sleep behaviour both affects and is affected by health conditions such as obesity, which causes obstructive sleep apnoea syndrome (Figorilli et al., 2025), chronic cardiometabolic conditions (Wang et al., 2023; Yin et al., 2017), depression (Liu, 2022), as well as behaviours such as physical activity. Estimating the incremental effect of sleep behaviour therefore requires model adjustment for these confounding factors and this

bidirectionality. A directed acyclic graph (DAG) was developed to formalise the assumed causal structure and guide covariate selection for adjustment. The DAG illustrates the interdependencies between sleep behaviour, health status, behavioural factors and healthcare costs (Figure 1).

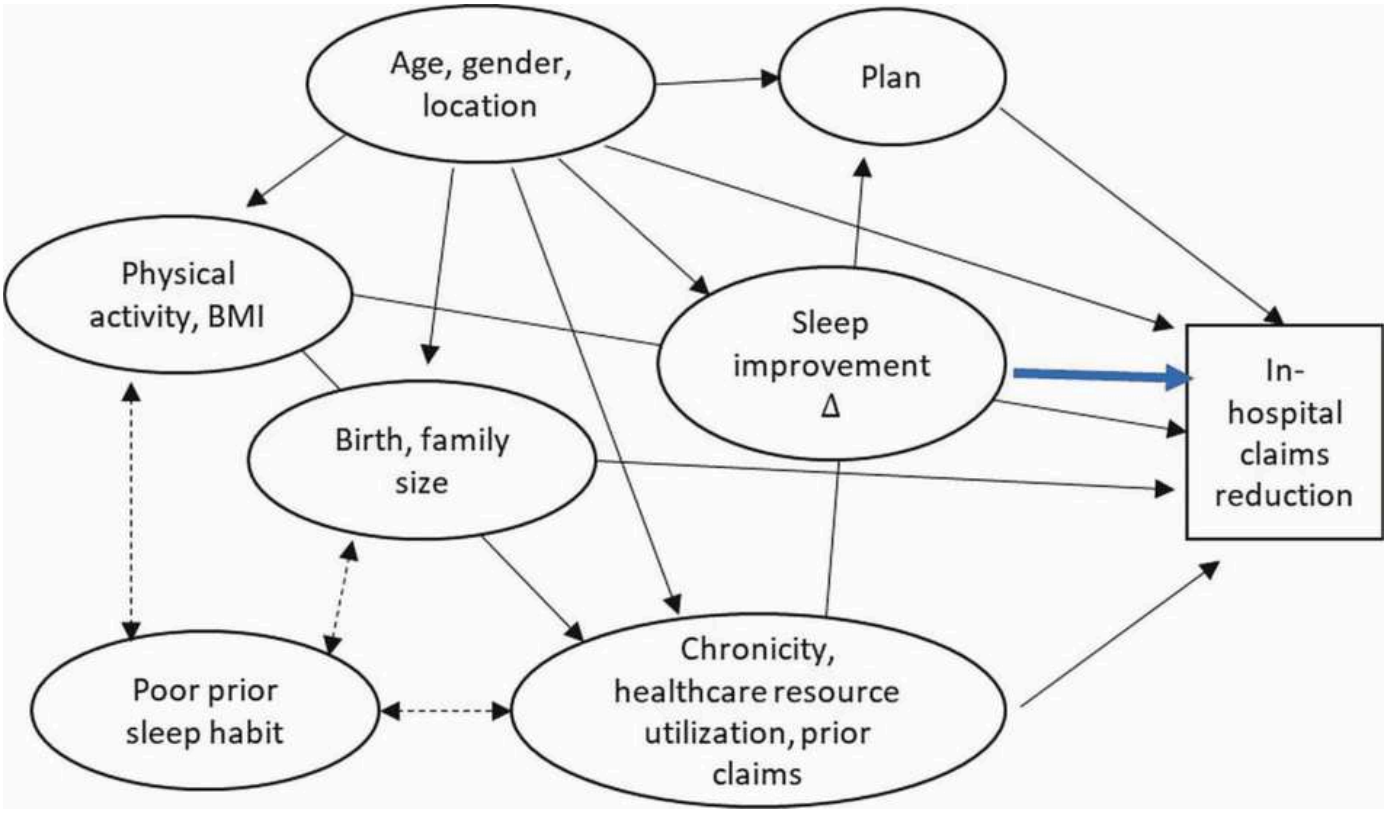

Figure 1: Directed acyclic graph of confounding factors affecting sleep and healthcare costs

Given the observational design of this study, members who improved their sleep behaviour may differ systematically from those who did not in ways that also influence in-hospital healthcare costs. Healthcare cost variation is likely to be strongly correlated with differences in morbidity, prior healthcare utilisation, medical scheme plan type and baseline sleep behaviour within the population. A direct comparison of raw costs between members with better and worse sleep would therefore conflate differences in baseline risk with any effect of a change in sleep behaviour.

The analysis therefore adjusted for a comprehensive set of risk factors measured prior to the sleep change (i.e., during the baseline period). These included demographic characteristics, underlying health status, chronic disease burden, prior hospital costs, medical scheme plan type, physical activity, and baseline sleep behaviour. Adjustment for these covariates was intended to improve comparability across exposure groups and reduce differences in underlying risk profile that could impact the measurement of the association between sleep behaviour and healthcare costs.

The risk-adjustment factors included in the analysis are summarised in Table 2 below.

Table 2: Sleep modelling risk-adjustment factors

Category	Risk-adjustment factors
Demographics	• Age • Sex • Body mass index (BMI)
Insurance plan details	• Medical scheme plan type
Location and family policy	• Postal code • Policy size
Health status	• Count of chronic conditions (and changes) • Resource Utilisation Band (RUB): a clinical risk score from 1 to 5 reflecting historical healthcare claims (Johns Hopkins ACG® system) • Presence of any diagnosed sleep disorder (insomnia or obstructive sleep apnoea) • Depression diagnosis • Recent childbirth (as a potential sleep disruptor)
Health behaviour	• Physical activity level (Vitality physical activity points in 2022)
Prior utilisation and sleep	• Baseline sleep habit score (to account for regression to the mean) • Prior-year (2021 or 2022) in-hospital cost

Descriptive characteristics

The final cohort comprised 28,808 adult medical scheme members with linked wearable and claims data across the study period (2022–2024). Baseline demographic, clinical and behavioural characteristics of the cohort are summarised in Table 3. These descriptive analyses provide context for interpreting the findings and assessing their generalisability to a broader medical scheme population.

Table 3: Baseline characteristics of the study cohort (2022)

Characteristic	Value (2022)
Age, mean (SD)	43 years (±12)
Male sex, %	55%
Chronic conditions, mean (SD)	0.6 (±0.8)
Proportion on Diamond Vitality status	40%
Device type: Garmin users	94%
Sleep duration, mean (SD)	7.6 h (±1.5)
Sleep regularity (bedtime variance)	1.1 h median nightly variation (IQR 0.6–1.3 h)
Deep sleep, mean (SD)	16.7% (±6.8) of sleep time
REM sleep, mean (SD)	19.2% (±7.3) of sleep time

Note: Values are presented as percentages or means (± standard deviation). Sleep regularity (bedtime variance) is defined as the median absolute deviation in sleep onset time from usual bedtime, measured in hours (lower values indicate a more consistent sleep schedule). IQR = interquartile range; REM = rapid eye movement; SD = standard deviation.

The cohort had a mean age of 43 years, compared to 46 years for the adult scheme population overall at that time, and 55% of those analysed were male. This cohort experienced relatively lower chronicity and multimorbidity, with only ~5% with >2 concurrent chronic conditions compared to 12% for the adult scheme population overall at that time. Forty percent of this cohort were on Vitality Diamond status, which is the highest level of engagement in the Vitality wellness programme, compared to 9% of the Vitality membership base on average. This subset of members was therefore generally healthier and more engaged in the wellness programme than the remaining medical scheme members: on average, they were younger, had fewer chronic conditions and recorded higher levels of physical activity.

The cohort of members slept for 7.6 hours per night on average (standard deviation of 1.5 hours), with substantial individual variation (interquartile range 6.7–8.6 hours). Many members displayed suboptimal sleep patterns initially: only approximately 18% were in the Good sleep habit band in 2022, versus 61% in the Medium habit band and 21% in the Poor habit band. Half of the cohort tracked their sleep almost daily; even so, only 11% of very frequent trackers managed to get both 7–9 hours of sleep and consistent sleep start times (within ±1 hour) at least 6 nights per week. Thus, there was scope for improvement in sleep behaviour and sleep habit band shift across the cohort.

Modelling approach

To examine the relationship of sleep, and of changes in sleep behaviour, with health outcomes, and because sleep improvement could not be experimentally assigned, we conducted a longitudinal observational study on a cohort of members with continuous medical scheme membership from 2022 to 2024 who tracked their sleep over the period. Members who changed their sleep behaviour from a Poor sleep habit band to a better band in the next year were used to determine the impact of habit formation on in-hospital healthcare costs relative to those who continued with poor sleep habits. Members with Poor or Medium sleep habits, whose sleep behaviour remained unchanged or worsened over the observation period, were assigned to the control group. Members who improved their sleep habit, or remained in a Good sleep habit, were assigned to the treatment group.

Two complementary statistical modelling approaches were implemented in Python to test whether the findings were directionally consistent. Both approaches used gradient-boosted trees to model complex relationships in the data.

1. An extreme gradient boosting (XGBoost) risk adjustment model was used to estimate expected 2024 in-hospital costs based on baseline demographic, clinical and utilisation characteristics measured in 2022, excluding sleep change variables. The difference between observed and expected costs was then compared to whether sleep habits improved, providing an estimate of cost differences associated with sleep improvement after adjustment for underlying risk. This approach adds value in isolating healthcare cost variation beyond what would be expected based on differing risk profiles, thus giving a more interpretable benchmark of relative cost performance for those who improved or maintained their sleep habits. This approach aligns with traditional actuarial risk-adjustment principles.

2. An X-learner ‘causal inference’ modelling package in Python was then used to estimate the average treatment effect of sleep improvement on hospital costs. This approach first determined how likely each person was to improve their sleep based on their starting characteristics and then compared outcomes between similar members whose sleep improved and those whose sleep did not. In comparison to traditional risk adjustment (1), this method more explicitly accounts for differences between members who did and did not improve their sleep, enabling a comparison between more closely matched groups and therefore a better-adjusted estimate of the average effect of improved sleep on in-hospital costs. The modelling package is well suited to real-world health data because it can account for complex, non-linear relationships between factors. This strengthens the interpretation of the observed differences as reflecting underlying associations rather than baseline differences in risk.

Together, these approaches allowed the association between improved sleep and subsequent in-hospital costs to be assessed using both predictive risk adjustment and causal treatment-effect estimation. We report the estimated percentage difference in mean in-hospital costs associated with improved sleep vs no improvement. Confidence intervals (95% CI) were derived in each approach to assess the statistical significance of each measure.

Ethical considerations

This study involved a secondary analysis of de-identified medical scheme membership data and did not involve any direct patient intervention. All data were anonymised prior to analysis and handled in accordance with applicable data protection and privacy regulations. The study was approved by the medical scheme's Research Governance Committee.

RESULTS

Sleep habit transitions

We analysed how the sleep behaviour of this population changed over the period, to evaluate the extent to which the membership exhibited a change in their sleep habits. The transitions between various sleep categories are shown in Table 4 below.

Table 4: Sleep habit band shift and maintenance in the study population

	Sleep habit at observation period		
Sleep habit at baseline period	Good	Medium	Poor
Good	62%	36%	2%
Medium	9%	80%	11%
Poor	1%	43%	56%

Note: Values are row percentages.

Among members classified in the Poor habit band in 2022, 43% of these members moved up to the Medium habit band in 2023. Movements of these members to the Good habit band were rare, with only ~1% of the Poor habit population improving all the way to Good. Sleep habits in the Medium habit band were more persistent: 80% remained Medium a year later, with 9% improving to Good and 11% worsening to Poor. Among those already with a Good sleep habit band in 2022, 62% sustained it in 2023, while 36% moved to Medium and 2% to Poor. Overall, these transitions indicate that meaningful behaviour change was achievable for a subset of previously Poor or Medium habit band sleepers, providing an opportunity to evaluate downstream health impacts.

Sleep habit change for each member was determined by any upward or downward movement from 2022 to 2023. The primary "treatment" of interest was an improved sleep habit band from baseline (Poor to Medium, Poor to Good, or Medium to Good). In addition, we separately examined maintenance in the Good sleep habit band (among those Good at baseline) as a further scenario reflecting the impact of sustained optimal sleep on healthcare costs.

Association of sleep habit changes with in-hospital health claims

Using the modelling framework described above, in-hospital claims costs in 2024 were compared between members who improved their sleep habit band and those who did not, after adjusting for the baseline risk factors. The model outputs are expressed as the estimated relative percentage difference in mean in-hospital claims costs (where a negative value indicates lower costs) with 95% confidence intervals, to indicate the statistical significance of these results. The results across both modelling approaches were compared to evaluate consistency in the direction and magnitude of observed associations.

Table 5 summarises the estimated risk-adjusted in-hospital costs for members who improved their sleep habit band compared to those who did not.

Table 5: Association of sleep habit band changes from 2022 to 2023 with relative difference in 2024 in-hospital health claims

Sleep habit change	Transition	X-learner model	Risk-adjustment model
Improvement in habit	Poor → Medium, Poor → Good or Medium → Good	−7.5% (95% CI: −17.3% to +2.3%)	−7.4% (95% CI: −14.2% to +0.1%)
	No improvement (reference)	0%	0%
Maintenance of habit	Good → Good	−5.2% (95% CI: −9.9% to −0.5%)	−4.9% (95% CI: −9.7% to −0.1%)
	No maintenance (reference)	0%	0%

Note: Negative percentages denote lower costs compared with the respective reference group. Cost outcomes are adjusted for baseline differences in demographics, chronic disease burden and prior utilisation.

In the risk-adjusted model, members who improved their sleep habit band from Poor/Medium to a higher habit band had 7.4% lower in-hospital claims costs in 2024 compared to similar members who did not improve (95% CI: −14.2% to +0.1%). Both modelling approaches yielded similar point estimates, but both 95% confidence intervals included zero, so the reduction in in-hospital costs associated with sleep habit improvement was not statistically significant.

Members who maintained a Good sleep habit displayed statistically significantly lower costs of approximately 5% in 2024 relative to those whose sleep habits deteriorated and moved from Good to a lower habit band. This estimate was more precise than that observed for sleep improvement: the X-learner's 95% CI for cost difference was approximately −9.9% to −0.5%, entirely below zero, indicating that sustaining optimal sleep patterns is associated with lower subsequent in-hospital costs.

Association between individual sleep metrics and in-hospital claims costs

Beyond the composite sleep habit measure, we quantified the relative contributions of specific sleep dimensions to reductions in in-hospital claims costs using the same methodological framework outlined above.

Table 6: Association of sleep metric changes from 2022 to 2023 with relative difference in 2024 in-hospital health claims

Sleep metric change	Relative difference in 2024 in-hospital claims (risk-adjustment model)
Improved sleep duration (<7 h to ≥7 h)	1.5% (95% CI: −10.8% to 7.6%)
Improved regularity (>1 h to ≤1 h variance)	−8.8% (95% CI: −17% to −0.7%)
Improved deep sleep (≤25% to >25%)	−1.9% (95% CI: −15.3% to 16.5%)

Although sample sizes for these specific metric changes were smaller, point estimate impacts for the 'Risk Adjustment' model suggested that improving sleep regularity (consistent sleep start time) had the largest potential effect, with an estimated 8.8% lower costs when individuals achieved <1 hour nightly variation versus remaining highly irregular (95% CI −17% to −0.7%). Increasing sleep duration to ≥7 hours per night (from <7 hours) was associated with an estimated ~1.5% lower cost (wide CI), while increasing deep sleep (to >25% of total sleep time) showed ~1.9% lower costs. Apart from regularity, these estimates were not statistically significant with current sample sizes, but they are consistent with the broader literature emphasising the importance of both adequacy and consistency of sleep, highlighting areas for further analysis as larger datasets become available.

DISCUSSION

This study provides novel evidence on the association between improvements in sleep behaviour and subsequent healthcare costs in a real-world adult population. Medical scheme members who improved their overall sleep habits had an estimated ~7% lower subsequent in-hospital claims costs (95% CI −14% to +0.1%) than similar members who did not improve their sleep habits. Lower healthcare costs were observed even after adjusting for a comprehensive set of baseline demographic, clinical and lifestyle factors, suggesting that better sleep may be independently associated with lower healthcare needs and therefore lower healthcare costs.

The consistency of findings using two analytical methods (a predictive risk adjustment and a causal inference model) strengthens the credibility of the findings, although some estimates (especially for partial improvements in individual sleep metrics) did not attain statistical significance with current sample sizes. To our knowledge, this study is among the first to quantify healthcare cost savings associated with better sleep in a cohort of medical scheme members. Prior research has established that individuals with sleep disorders incur higher healthcare expenditures (Olatunde & Patton, 2025; Wickwire et al., 2019). However, evidence linking behavioural improvements in sleep to lower healthcare costs at scale is scant. Our results align with clinical intuition and extend findings from controlled settings (de Gans et al., 2025). Our population-level analysis suggests that even in free-living conditions, those who improve their sleep habits may experience health benefits large enough to lower their need for hospital-based care. Given the relatively low cost and broad applicability of sleep interventions, even modest reductions in healthcare utilisation could yield favourable population-level cost-effectiveness.

LIMITATIONS

These results are from observational data and cannot definitively prove causation. Despite extensive covariate adjustment, residual confounding may remain. For instance, individuals who improved their sleep might also have improved other health behaviours or had unmeasured favourable characteristics (e.g., greater self-efficacy or social support) that contributed to lower in-hospital healthcare use. We attempted to mitigate this by controlling for many proxies of health status and behaviour (including prior physical activity, chronic disease changes and policy size), but unmeasured factors could bias results.

The selection bias inherent in our population is a limitation: participants were all members of a medical scheme plan with wearable devices, with higher recorded physical activity and lower chronic condition count than the scheme membership at large. Thus, baseline hospital utilisation in this group was already relatively low, and the absolute reductions observed might differ in other populations, for example, people not on a private medical plan.

Our focus on in-hospital claims costs means we did not examine potential shifts in outpatient or pharmacy utilisation; it is possible that improved sleepers had fewer hospitalisations but more primary care engagement.

Measurement limitations must also be acknowledged. Sleep was tracked via consumer wearable devices, not clinical polysomnography (PSG). Research comparing gold-standard PSG with commercially available sleep trackers suggests these devices are relatively accurate in identifying sleep onset and offset, allowing reliable estimation of sleep duration. Finger-worn trackers, such as the Oura Ring, demonstrate greater than 95% sensitivity compared with PSG (de Zambotti et al., 2019). While these devices provide reasonable estimates of sleep duration and regularity, their measures of sleep stages require further validation (de Zambotti et al., 2024).

We partially addressed this by focusing on the composite Sleep Habit Score and broad categories of improvement, rather than precise changes in minutes of REM sleep. Nonetheless, some misclassification in sleep metrics is likely, which would generally bias towards the null (making it harder to detect true effects). Finally, our confidence intervals for some estimates were wide, reflecting the limited number of individuals who achieved substantial sleep improvements within the study window. For example, only ~1,000 members lengthened their average sleep by ≥1 hour (to reach ≥7 hours nightly), hampering statistical power for further analyses, such as of mortality. As the size of this sleep-tracking cohort grows over time, future analyses will be able to estimate these effects with greater precision and confidence.

POLICY IMPLICATIONS

Our findings, together with existing evidence on sleep's health importance, carry several implications for health policy and preventive care strategies:

- Integrate sleep health into chronic disease care: Routinely screen and manage sleep disorders in patients with chronic conditions such as diabetes, hypertension, and depression, as improving sleep may enhance outcomes and reduce complications and healthcare costs.
- Promote healthy sleep behaviours: Encourage sleep improvement through wellness programmes, insurance incentives, and public education, as better sleep health could lower healthcare claims and support large-scale behaviour change.
- Expand coverage for sleep interventions: Consider broader funding for evidence-based sleep therapies, including sleep apnoea diagnosis and CPAP treatment, as well as cognitive behavioural therapy for insomnia, to reduce long-term healthcare costs and complications.

Ultimately, it is important to recognise that sleep is not only a primordial need but also a key behavioural intervention for better health, alongside healthy nutrition and physical activity. This study adds to a growing body of research indicating that investments in sleep health can pay dividends in terms of lower healthcare burden and improved population wellbeing. Further research, including randomised trials of sleep improvement programmes, will help solidify causal linkages and guide implementation of sleep interventions within health systems.

CONCLUSION

In this large observational study, improved sleep habits were associated with lower subsequent in-hospital healthcare costs among adult medical scheme members. The findings suggest that in a real-world, non-clinical setting, individuals who achieve better sleep habits may experience meaningfully lower in-hospital costs. These results strengthen the case for incorporating sleep promotion into public health and preventive care strategies. However, due to the non-randomised design and relatively health-conscious study population, the results should be interpreted with caution and may not be generalisable to other populations. Future research should aim to confirm these associations in more diverse populations and through controlled interventions. If causal, the magnitude of in-hospital cost savings observed indicates that population-level improvements in sleep could have a significant impact on healthcare demand and expenditures, alongside the clear benefits for individual wellbeing.

REFERENCE LIST

Antza, C., Kostopoulos, G., Mostafa, S., Nirantharakumar, K., & Tahrani, A. (2021). The links between sleep duration, obesity and type 2 diabetes mellitus. *Journal of Endocrinology*, *252*(2), 125–141. https://doi.org/10.1530/JOE-21-0155

Bock, J. M., Vungarala, S., Covassin, N., & Somers, V. K. (2022). Sleep duration and hypertension: Epidemiological evidence and underlying mechanisms. *American Journal of Hypertension*, *35*(1), 3–11. https://doi.org/10.1093/ajh/hpab146

Cappuccio, F. P., D'Elia, L., Strazzullo, P., & Miller, M. A. (2010). Sleep duration and all-cause mortality: A systematic review and meta-analysis of prospective studies. *Sleep*, *33*(5), 585–592. https://doi.org/10.1093/sleep/33.5.585

de Gans, C. J., Meewisse, A. J. G., van den Ende, E. S., van Zuylen, M. L., Nanayakkara, P. W. B., Hermanides, J., & Stenvers, D. J. (2025). The effects of sleep-improving interventions in medical hospital wards: The WEsleep study—A randomized clinical trial. *European Journal of Internal Medicine*, *139*, Article 106308. https://doi.org/10.1016/j.ejim.2025.04.015

de Zambotti, M., Goldstein, C., Cook, J., Menghini, L., Altini, M., Cheng, P., & Robillard, R. (2024). State of the science and recommendations for using wearable technology in sleep and circadian research. *Sleep*, *47*(4), Article zsad325. https://doi.org/10.1093/sleep/zsad325

de Zambotti, M., Rosas, L., Colrain, I. M., & Baker, F. C. (2019). The sleep of the ring: Comparison of the oURA sleep tracker against polysomnography. *Behavioral Sleep Medicine*, *17*(2), 124–136. https://doi.org/10.1080/15402002.2017.1300587

Figorilli, M., Velluzzi, F., & Redolfi, S. (2025). Obesity and sleep disorders: A bidirectional relationship. *Nutrition, Metabolism and Cardiovascular Diseases*, *35*(6), Article 104014. https://doi.org/10.1016/j.numecd.2025.104014

Giuntella, O., & Mazzonna, F. (2019). Sunset time and the economic effects of social jetlag: Evidence from US time zone borders. *Journal of Health Economics*, *65*, 210–226. https://doi.org/10.1016/j.jhealeco.2019.03.007

Grandner, M., Mullington, J. M., Hashmi, S. D., Redeker, N. S., Watson, N. F., & Morgenthaler, T. I. (2018). Sleep duration and hypertension: Analysis of >700,000 adults by age and sex. *Journal of Clinical Sleep Medicine*, *14*(6), 1031–1039. https://doi.org/10.5664/jcsm.7176

Grandner, M. A., Seixas, A., Shetty, S., & Shenoy, S. (2016). Sleep duration and diabetes risk: Population trends and potential mechanisms. *Current Diabetes Reports*, *16*(11), Article 106. https://doi.org/10.1007/s11892-016-0805-8

Hosseini, K., Soleimani, H., Tavakoli, K., Maghsoudi, M., Heydari, N., Farahvash, Y., Etemadi, A., Najafi, K., Askari, M. K., Gupta, R., Hakim, D., & Rahimi, K. (2024). Association between sleep duration and hypertension incidence: Systematic review and meta-analysis of cohort studies. *PLOS ONE*, *19*(7), Article e0307120. https://doi.org/10.1371/journal.pone.0307120

Hui, S. K. A., & Grandner, M. A. (2015). Trouble sleeping associated with lower work performance and greater health care costs: Longitudinal data from Kansas State employee wellness program. *Journal of Occupational and Environmental Medicine*, *57*(10), 1031–1038. https://doi.org/10.1097/JOM.0000000000000534

Huyett, P., & Bhattacharyya, N. (2021). Incremental health care utilization and expenditures for sleep disorders in the United States. *Journal of Clinical Sleep Medicine*, *17*(10), 1981–1986. https://doi.org/10.5664/jcsm.9392

Irwin, M. R. (2015). Why sleep is important for health: A psychoneuroimmunology perspective. *Annual Review of Psychology*, *66*, 143–172. https://doi.org/10.1146/annurev-psych-010213-115205

Johns Hopkins University. (n.d.). *The Johns Hopkins ACG® System*. https://www.hopkinsacg.org

Kim, S. W., & Madhavan, S. (2024). Employment, income, and sleep in South Africa: An actor-partner interdependence model approach. *Sleep Health*, *10*(6), 643–650. https://doi.org/10.1016/j.sleh.2024.07.006

Liu, Y., Wheaton, A. G., Chapman, D. P., Cunningham, T. J., Lu, H., & Croft, J. B. (2016). Prevalence of healthy sleep duration among adults —United States, 2014. *Morbidity and Mortality Weekly Report*, *65*(6), 137–141. https://doi.org/10.15585/mmwr.mm6506a1

Lloyd-Jones, D. M., Allen, N. B., Anderson, C. A. M., Black, T., Brewer, L. C., Foraker, R. E., Grandner, M. A., Lavretsky, H., Perak, A. M., Sharma, G., Rosamond, W., & American Heart Association. (2022). Life's Essential 8: Updating and enhancing the American Heart Association's construct of cardiovascular health: A presidential advisory from the American Heart Association. *Circulation*, *146*(5), e18–e43. https://doi.org/10.1161/CIR.0000000000001078

Lu, H., Yang, Q., Tian, F., Lyu, Y., He, H., Xin, X., & Zheng, X. (2021). A meta-analysis of a cohort study on the association between sleep duration and type 2 diabetes mellitus. *Journal of Diabetes Research*, *2021*, Article 8861038. https://doi.org/10.1155/2021/8861038

National Sleep Foundation. (2025). *Sleep in America® poll—2025: Key findings*. https://www.thensf.org/sleep-in-america-polls/

Olatunde, K., & Patton, S. (2025). Association between insomnia and healthcare utilization: A scoping review of the literature. *American Journal of Lifestyle Medicine, 19*(3), 403–418. https://doi.org/10.1177/15598276231164953

Philip, P., Chaufton, C., Orriols, L., Lagarde, E., Amoros, E., Laumon, B., Åkerstedt, T., Taillard, J., & Sagaspe, P. (2014). Complaints of poor sleep and risk of traffic accidents: A population-based case-control study. *PLOS ONE*, *9*(12), Article e114102. https://doi.org/10.1371/journal.pone.0114102

Ramar, K., Malhotra, R. K., Carden, K. A., Martin, J. L., Abbasi-Feinberg, F., Aurora, R. N., Kapur, V. K., Olson, E. J., Rosen, C. L., Rowley, J. A., Shelgikar, A. V., & Trotti, L. M. (2021). Sleep is essential to health: An American Academy of Sleep Medicine position statement. *Journal of Clinical Sleep Medicine*, *17*(10), 2115–2119. https://doi.org/10.5664/jcsm.9476

Robbins, R., & Quan, S. (2024). Sleep health. *NEJM Evidence*, *3*(8), Article EVIDra2300269. https://doi.org/10.1056/EVIDra2300269

Shah, A. S., Pant, M. R., Bommasamudram, T., Nayak, K. R., Roberts, S. S. H., Gallagher, C., Vaishali, K., Edwards, B. J., Tod, D., Davis, F., & Pullinger, S. A. (2025). Effects of sleep deprivation on physical and mental health outcomes: An umbrella review. *American Journal of Lifestyle Medicine*. Advance online publication. https://doi.org/10.1177/15598276251346752

St-Onge, M.-P., Aggarwal, B., Fernandez-Mendoza, J., Johnson, D., Kline, C. E., Knutson, K. L., Redeker, N., & Grandner, M. A. (2025). Multidimensional sleep health: Definitions and implications for cardiometabolic health: A scientific statement from the American Heart Association. *Circulation: Cardiovascular Quality and Outcomes*, Article e000139. https://doi.org/10.1161/HCQ.0000000000000139

Wang, C., Liu, Y., Chen, X., Zhu, J., Wu, Q., Chen, H., Liao, H., Lin, J., Wang, Z., Zheng, Z., & Chen, R. (2023). Meta-analysis of correlation between sleep duration and gender difference in adults with type 2 diabetes. *Sleep and Breathing*, *27*(6), 2325–2332. https://doi.org/10.1007/s11325-023-02841-0

Watson, N. F., Badr, M. S., Belenky, G., Bliwise, D. L., Buxton, O. M., Buysse, D., Dinges, D. F., Gangwisch, J., Grandner, M. A., Kushida, C., Malhotra, R. K., Martin, J. L., Patel, S. R., Quan, S. F., & Tasali, E. (2015). Joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society on the recommended amount of sleep for a healthy adult: Methodology and discussion. *Journal of Clinical Sleep Medicine*, *11*(8), 931–952. https://doi.org/10.5664/jcsm.4950

Wickwire, E. M., Tom, S. E., Scharf, S. M., Vadlamani, A., Bulatao, I. G., & Albrecht, J. S. (2019). Untreated insomnia increases all-cause health care utilization and costs among Medicare beneficiaries. *Sleep*, *42*(4), Article zsz007. https://doi.org/10.1093/sleep/zsz007

Windred, D. P., Burns, A. C., Lane, J. M., Saxena, R., Rutter, M. K., Cain, S. W., & Phillips, A. J. K. (2024). Sleep regularity is a stronger predictor of mortality risk than sleep duration: A prospective cohort study. *Sleep*, *47*(1), Article zsad253. https://doi.org/10.1093/sleep/zsad253

Yin, J., Jin, X., Shan, Z., Li, S., Huang, H., Li, P., Peng, X., Peng, Z., Yu, K., Bao, W., Yang, W., Chen, X., & Liu, L. (2017). Relationship of sleep duration with all-cause mortality and cardiovascular events: A systematic review and dose-response meta-analysis of prospective cohort studies. *Journal of the American Heart Association*, *6*(9), Article e005947. https://doi.org/10.1161/JAHA.117.005947

DECLARATIONS

Author contributions

Katherine Tregoning: Formal Analysis Methodology, Writing, Review & Editing; Simon Henrie: Conceptualising Study, Formal Analysis, Writing and Reviewing. Cassi Rubin: Data Curation, Writing, Editing. Deepak Patel: Writing, Validation, Review & Editing. Lara Wayburne: Conceptualising, Overseeing analysis, Writing and Editing. Claudia Pastellides: Analysis, Methodology, Initial Writing. Roseanne Harris: Reviewing and editing. Mosima Mabunda: Reviewing and editing.

All authors have read and approved the final version of this manuscript.

Data availability statement

The data supporting the findings of this study are available from the authors upon reasonable request, subject to ethical approval and institutional regulations.

Declaration of competing interest

The authors are employees of Discovery Health and Discovery Vitality (see Conflict of interest below).

Conflict of interest

The authors are employees of Discovery Health and Discovery Vitality, who provided anonymised data for this study. Discovery Health and Discovery Vitality offer a wellness programme for sleep to their full membership. The authors have aimed to design and interpret the study without bias, but readers should be aware of this potential conflict of interest. All analyses were conducted in an actuarially robust and data-driven manner, and we report findings irrespective of their alignment with commercial interests.

Declaration of assistive technologies in the writing process

The authors declare that generative artificial intelligence (GenAI) and AI-assisted tools were used in a limited and responsible manner during the preparation of this manuscript. Specifically, Microsoft Copilot was used for language refinement and structuring of sections, and validated by the authors to ensure accuracy, originality, and compliance with ethical and scholarly standards. The authors take full responsibility for the final content of this manuscript.



STREAM IV

DIGITAL HEALTH AND DATA-DRIVEN PREVENTION

ASSOCIATION BETWEEN RECORDED PHYSICAL ACTIVITY AND CANCER PROGRESSION OR MORTALITY IN INDIVIDUALS DIAGNOSED WITH CANCER IN SOUTH AFRICA

Ntokozo Mabena, Nivash Rugbeer, Sandra Lehmann, Georgia Torres, Deepak Patel, Mosima Mabunda, Mike Greyling, Jane S Thornton, Yun-Hee Choi, Saverio Stranges and Jon S Patricios

British Journal of Sports Medicine, 2025, *59*(10), 715–721. https://doi.org/10.1136/bjsports-2024-108813. Reprinted under the CC BY-NC 4.0 licence; abstract headings added for consistency with this issue. https://pubmed.ncbi.nlm.nih.gov/39773836/

ABSTRACT

Background: This study aimed to determine the association between progression and mortality in individuals with stage 1 cancer and their recorded physical activity before the diagnosis of the cancer.

Methods: We included 28,248 members with stage 1 cancers enrolled in an oncology programme in South Africa. Physical activity was recorded using fitness devices, logged gym sessions and participation in organised fitness events. Levels of physical activity over the 12 months before cancer diagnosis were categorised as no physical activity, low physical activity (an average of <60 min/week) and moderate to high physical activity (≥60 min/week). Measured outcomes were time to progression, time to death and all-cause mortality.

Results: Physically active members showed lower rates of cancer progression and lower rates of death from all causes. The HR for progression to higher stages or death was 0.84 (95% CI 0.79 to 0.89), comparing low activity with no physical activity, and 0.73 (95% CI 0.70 to 0.77), comparing medium to high physical activity with no physical activity. The HR for all-cause mortality was 0.67 (95% CI 0.61 to 0.74), comparing low physical activity with no activity, and 0.53 (95% CI 0.50 to 0.58), comparing medium to high physical activity with no physical activity.

Conclusions: Individuals engaging in any level of recorded physical activity showed a reduced risk of cancer progression or mortality compared with those not physically active. There was a further reduction among individuals with moderate to high levels of physical activity compared with those with lower levels.

ACKNOWLEDGEMENT OF THE RESEARCH SYMPOSIUM

2026
CMS RESEARCH SYMPOSIUM









The successful hosting of the CMS Research Symposium 2026 would not have been possible without the valuable contributions and support of various individuals and stakeholders.











We hope that the insights and knowledge shared during this symposium will inspire continued scholarly engagement, collaborative research and practical solutions to societal challenges.

ACKNOWLEDGEMENT OF INDABA PARTNERS

The Council for Medical Schemes gratefully acknowledges the partners of the CMS Industry Indaba 2026, whose support made the Indaba possible (see the partner logos below):

- South African Police Service Medical Scheme (POLMED)
- Government Employees Medical Scheme (GEMS)
- Insight Actuaries
- Public Service Co-ordinating Bargaining Council (PSCBC)
- Barnard Medical
- hearConnect

The Council also thanks the organisations whose leaders contributed to the programme as keynote speakers, presenters, panellists and moderators:

- National Department of Health
- Health Market Inquiry (former Chair)
- Board of Healthcare Funders
- Health Funders Association
- Office of Health Standards Compliance
- Health Professions Council of South Africa
- South African Medical Association (SAMA)
- National Health Care Professionals Association (NHCPA)
- Discovery Health
- Medscheme
- Alchemy Health Technologies
- MediCheck
- Solutionist Thinkers
- South African Medical Research Council (SAMRC)/University of the Witwatersrand Centre for Health Economics
- Stellenbosch University

The Council further extends its gratitude to all speakers, partners, exhibitors and delegates for their valued contribution to the success of the 2026 CMS Industry Indaba.



Source: CMS Industry Indaba 2026 partners page (Council for Medical Schemes, 2026). Logos are reproduced with acknowledgement of the respective organisations.