DVA Health Programs Evaluation Approaches Review

Final

Version control

|  |  |
| --- | --- |
| Document version | 3.0 |
| Date | September 2023 |
| Document status | Final |
| Author | Ernst & Young |
| Owner | Program Governance & Evaluation |
| File name | DVA Health Programs Evaluations Approaches Review Report – External |

Before use, please verify this document is current.

Disclaimer

This report was prepared at the request of The Australian Government Department of Veterans’ Affairs, Program Governance & Evaluation, solely for the purposes of identifying and reviewing relevant evaluation approaches and is not appropriate for use for other purposes. This report is provided for information purposes only in order to assist The Australian Government Department of Veterans’ Affairs, Program Governance & Evaluation in building an understanding of better practice evaluation approaches and should not be taken as providing specific advice on any issue, nor may this report be relied upon by any party other than The Australian Government Department of Veterans’ Affairs, Program Governance & Evaluation. In carrying out our work and preparing this report EY has worked solely on the instructions and information of The Australian Government Department of Veterans’ Affairs, Program Governance & Evaluation, and has not taken into account the interests or individual circumstances of any party other than The Australian Government Department of Veterans’ Affairs, Program Governance & Evaluation. EY does not accept any responsibility for use of the information contained in this report and makes no guarantee nor accept any legal liability whatsoever arising from or connected to the accuracy, reliability, currency or completeness of any material contained herein. EY expressly disclaims all liability for any costs, loss, damage, injury or other consequence which may arise directly or indirectly from use of, or reliance on, this report.

Contents

[Executive summary 4](#_Toc110956273)

[Review questions 4](#_Toc110956274)

[Review methodology 4](#_Toc110956275)

[Key Findings 4](#_Toc110956276)

[Conclusion 5](#_Toc110956277)

[1. Introduction 7](#_Toc110956278)

[1.1. Background 7](#_Toc110956279)

[1.2. Aims and objectives of the review 8](#_Toc110956280)

[2. Review Approach 9](#_Toc110956281)

[2.1. Desktop literature review 9](#_Toc110956282)

[2.1.1. Published literature review 9](#_Toc110956283)

[2.1.2. Grey literature search 11](#_Toc110956284)

[2.2. Key informant interviews 11](#_Toc110956285)

[2.3. Case studies 11](#_Toc110956286)

[2.4. Review search results 12](#_Toc110956287)

[3. Key findings 15](#_Toc110956288)

[3.1. What is better practice evaluation of Australian health programs? 15](#_Toc110956289)

[3.1.1. The right foundations are used 16](#_Toc110956290)

[3.1.2. A systematic approach is used when prioritising programs for evaluation 21](#_Toc110956291)

[3.1.3. Evaluation type is selected based on need or intended goals 22](#_Toc110956292)

[3.1.4. Careful consideration is given when choosing between internal and external evaluators 27](#_Toc110956293)

[3.1.5. Evaluation findings are used to support decision-making 28](#_Toc110956294)

[3.2. What types of evaluations are used for the assessment of health programs across Australia? 28](#_Toc110956295)

[3.3. How have health programs been evaluated within government agencies? Does this differ from the private sector? 30](#_Toc110956296)

[3.4. What is the cost in conducting these health program evaluations? 31](#_Toc110956297)

[3.5. Have the methods of health program evaluations changed over time? 32](#_Toc110956298)

[4. Considerations for better practice evaluation 34](#_Toc110956299)

[5. References 36](#_Toc110956300)

[Appendix A Desktop literature review summary 38](#_Toc110956301)

[Appendix B Key informant interviews summaries 49](#_Toc110956302)

[Appendix C Case study summaries 51](#_Toc110956303)

Executive summary

In April 2022, the Australian Government Department of Veterans’ Affairs (DVA) engaged EY to deliver research services (‘the review’) on better practice approaches in the evaluation of health and wellbeing programs (‘health programs’). The review aims to provide recommendations regarding the existing types of evaluation occurring within health programs across Australia (e.g. process, impact, economic/value for money, fidelity, outcome), including the appropriate use of each type of evaluation. These recommendations will help inform how DVA can plan for and carry out evaluations to support the wellbeing of veterans and families. This report presents the findings of the review.

Review questions

The review answers the following questions to help guide DVA on their journey to better practice evaluations:

1. *What is better practice evaluation of Australian health programs?*
2. *What types of evaluations are used for the assessment of health programs across Australia?*
3. *How have health programs been evaluated within government agencies? Does this differ from the private sector?*
4. *What is the cost in conducting these health program evaluations?*
5. *Have the methods of health program evaluations changed over time?*

Review methodology

The review methodology consisted of three data collection activities:

1. *Desktop literature review* to identify the Australian health program evaluation experience. This involved searching the published and grey literature for information on the types of evaluation occurring on health programs across Australia (e.g. process, impact, economic/value for money, fidelity, outcome). A total of 45 publications were identified – noting the relatively small number of health programs directly relevant to those offered by DVA.
2. *Key informant interviews* with two subject matter experts from the United States of America (US) to explore health evaluation practices in the US, which has a large number of health services and programs available to veterans.
3. *Case studies* on three Australian government agencies with experience in evaluation to understand evaluation practices of Australian organisations with mature established capability.

Key Findings

What is better practice evaluation of Australian health programs?

Five main principles characterise best practice evaluations:

1. The right foundations are used, including:
   1. embedding evaluation into program design with appropriate governance and oversight
   2. designing a program logic and evaluation framework for each program
   3. creating an evaluation culture within an organisation which allows for capacity building, adequate resource allocation, and general understanding of the benefits and importance of evaluation.
2. A systematic approach is used when prioritising programs for evaluation.
3. Choosing the appropriate evaluation type based on need or intended goals, while recognising that process, outcome and economic evaluations were the most common types of evaluation identified by policy and guidance documents from this review.
4. Careful consideration be given to choosing between internal or external evaluators, with external evaluators being useful for larger scale or high investment projects where independence is desirable.
5. Evaluation findings are used to support decision making and are usually published in order to inform policy and program development, as well as embed an evaluation culture within the organisation.

What types of evaluations are used for the assessment of health programs across Australia?

There is a large variety of types of health program evaluations; the most commonly found are process evaluations using mixed methods data collection strategies. Pre-post designs were commonly reported by the systematic reviews.

How have health programs been evaluated within government agencies? Does this differ from the private sector?

Australian government agencies commonly outsource their evaluations, whereas the private sector appears to both outsource evaluations and perform them in-house. There were otherwise no meaningful differences noted in evaluation types or processes between the government and private sector.

What is the cost in conducting these health program evaluations?

Guidance into costing of evaluations is rare and ambiguous. Numerous factors contribute to diversity in costs, including complexity in obtaining data, level of risk, maturity of program, internal vs external evaluators, intended use and significance to stakeholders and health outcomes. The NSW government has broadly suggested evaluations should be valued at 10% of program costs, but this figure can under- and overestimate potential health program costs. Costs should be informed by program and evaluation characteristics.

Have the methods of health program evaluations changed over time?

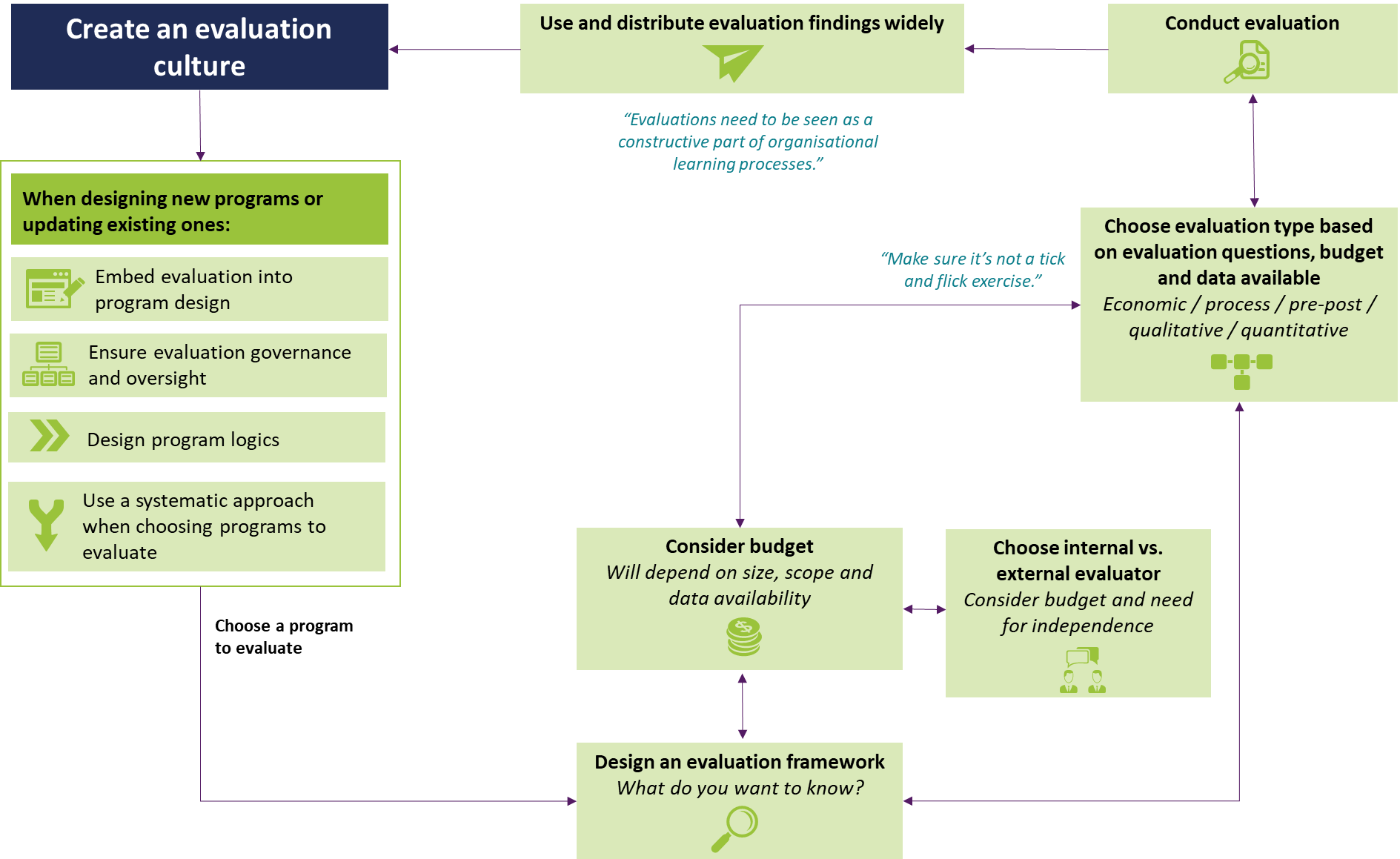
Health program evaluations are increasingly becoming embedded into program design rather than serving as an external function. Continuous evaluation is becoming more common than single one-off evaluations. There is evidence from overseas that stakeholder engagement is increasing in evaluations, both in informing program design, and in collecting qualitative data on attitudes and behaviour. Evaluation remains an important but largely undervalued part of health programs.

Conclusion

This review did not identify any firm parameters to inform decisions on evaluation design, budget, and the choice between using external versus internal evaluators. Although process evaluations were the most common of evaluation in the literature, this may have been due to budget constraints or the maturity of the program being evaluated rather than better practice per se. While it was found that government agencies mostly outsourced their evaluations; this may be a result of factors such as internal resourcing constraints and the time-limited nature of evaluations, rather than as a result of better practice.

Better practice evaluations were found to be associated with organisations with an evaluation culture – whereby evaluative thinking and a continuous improvement approach is embedded into organisations, leading to high quality evaluations and thus high-quality program design and delivery. This is articulated in Figure ES1.

Figure ES1: Creating an evaluation culture



1. Introduction

In April 2022, the Australian Government Department of Veterans’ Affairs (DVA) engaged EY to deliver research services (‘the review’) on better practice approaches in the evaluation of health and wellbeing programs (‘health programs’). The review aims to provide recommendations regarding the existing types of evaluation occurring within health programs across Australia (e.g. process, impact, economic/value for money, fidelity, outcome), including the appropriate use of each type of evaluation. This report presents the findings of the review.

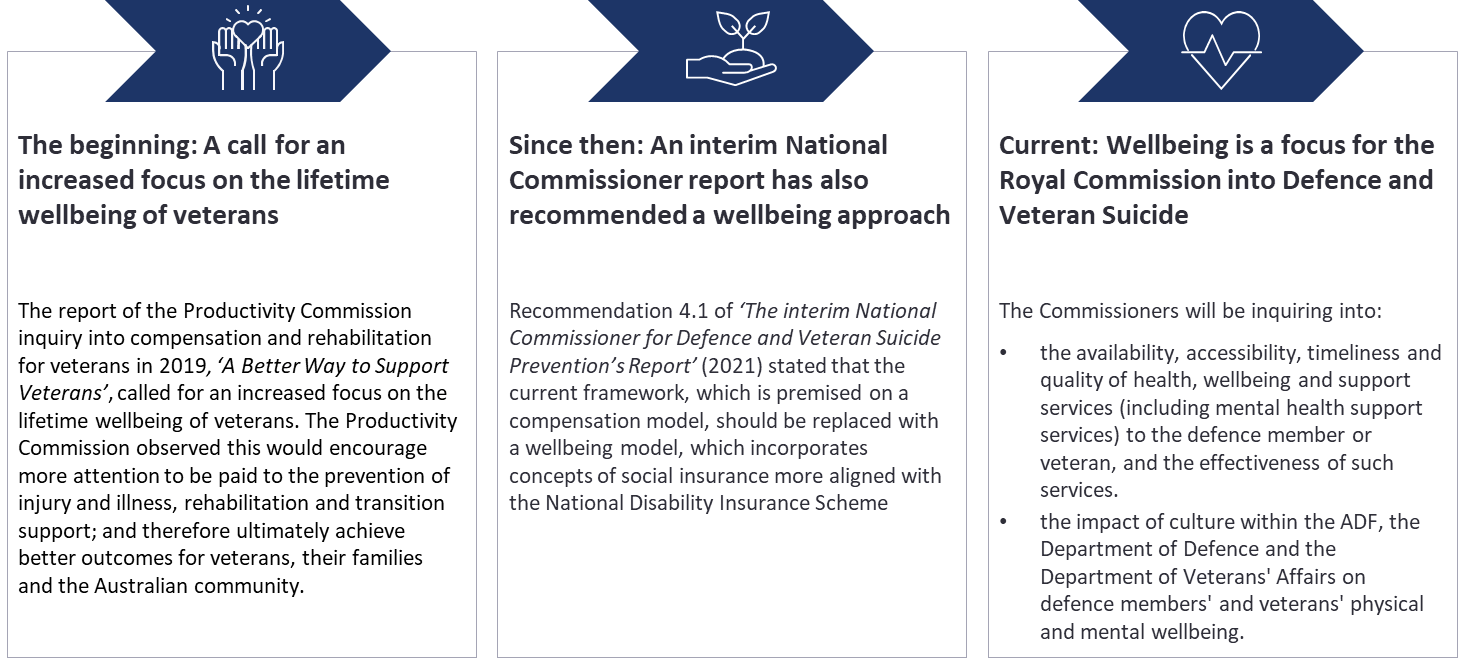
* 1. Background

The purpose of DVA is to support the wellbeing of those who serve or have served in the defence of our nation and their families, by:

* Partnering with organisations and individuals to help design, implement and deliver effective programs and benefits, which enhance the wellbeing of veterans and families; and
* Providing and maintaining war graves and delivering meaningful commemorative activities to promote community recognition and understanding of the service and sacrifice of veterans.

Since 2019, DVA has been on an evolving strategic journey to better live this purpose as summarised below (Figure 1).

Figure 1. The DVA strategic journey



To support this strategic intent, DVA has taken a proactive approach to strengthening its evidence-based programs which are aimed at enhancing the health and wellbeing needs of veterans and families. Evaluation enables the measurement of impact of these programs using quantitative and qualitative data and research, allowing for more robust decision-making and embedding best practice at the heart of delivery. As such, DVA has recognised the importance that evaluation has in shaping Departmental policy, programs and initiatives.

The DVA Program Governance & Evaluation section (PG&E) is responsible for the evaluation of DVA’s health and wellbeing programs. In line with DVA’s Strategic Research Framework 2019-21, PG&E seeks to develop a comprehensive evaluation strategy to ensure that gaps in DVA knowledge base are bridged, and that DVA may respond to the many multi-faceted and complex needs of DVA’s core clients – veterans and families.

To this end, PG&E is actively invested in developing internal capability to undertake health program evaluations and is currently responsible for an evaluation strategy and schedule to prioritise the type and timing of evaluation of DVA’s health programs.

* 1. Aims and objectives of the review

The review aims to detail the types of evaluation that are used to evaluate health programs in Australia. It will outline the strengths and weaknesses and appropriate use of each type of evaluation and provide an understanding of better practice evaluation of health programs. The outcome of the review will inform advice/recommendations to DVA on:

* an evidence-based, better practice approach for future health program evaluations in DVA; and
* opportunities to align DVA health program evaluations with better practice to deliver quality, evidence-based services that directly address the wellbeing needs of veteran community.

Results from this literature review will also complement work being conducted across the Department to build and develop DVA’s broader evaluation capability.

The review answers the following questions to guide DVA on their journey to better practice evaluations:

1. *What is better practice evaluation of Australian health programs?*
2. *What types of evaluations are used for the assessment of health programs across Australia?*
3. *How have health programs been evaluated within government agencies? Does this differ from the private sector?*
4. *What is the cost in conducting these health program evaluations?*
5. *Have the methods of health program evaluations changed over time?*

Together with the implementation of DVA’s Program Management Framework (PMF), this work will:

* support the improved governance of DVA’s health programs
* inform better program delivery and help staff deliver program outcomes
* improve DVA’s ability to deliver services to veterans and families that align with DVA’s current strategic priorities, enhance governance and accountability
* include a focus on wellbeing rather than illness.

1. Review Approach

This chapter outlines the approach and methodology to the review, which consisted of three data collection activities:

1. Desktop literature review – to identify the Australian health program evaluation experience
2. Key informant interviews – to explore comparable health evaluation practices abroad
3. Case studies – to understand evaluation practices of Australian organisations with mature established capability
   1. Desktop literature review

The desktop literature review consists of two components:

1. A published literature search, to understand the Australian experience and characterise the types of evaluations being conducted in Australia
2. A grey literature search to identify supporting information and guidance, in publicly available information on Australian evaluation practice
   * 1. Published literature review

The literature was searched for information on types of evaluation occurring within health programs across Australia (e.g. process, impact, economic/value for money, fidelity, outcome), including the appropriate use of each type of evaluation. The following strategy was used to conduct this search:

1. The following databases were searched using combinations of search terms outlined in Table 1 to generate initial results:

* PubMed (using the PubMed Advanced Search Builder and MeSH terms where relevant)
* CINAHL (using CINAHL subject headings where relevant)
* Google Scholar (using the Advanced Search function).

**Table 1. Primary, Secondary and Tertiary\* search terms, mapped against Key Review Questions**

| **Primary search term** | **Secondary search terms** | **Tertiary search terms** | **Evaluation Question** | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Q1** | **Q2** | **Q3** | **Q4** | **Q5** |
| Program Evaluation | * Health, healthcare, wellbeing * DVA, rehabilitation, veterans, defence | Methods, approaches, guidelines | 🗸 | 🗸 |  |  | 🗸 |
| Developmental, process, formative, outcomes, summative, impact | 🗸 | 🗸 |  |  | 🗸 |
| Cost Benefit Analysis, Return on Investment | 🗸 |  |  | 🗸 | 🗸 |
| Implementation, translation, fidelity |  |  |  |  | 🗸 |
| Theoretical frameworks, program theory, program logic, evaluative frameworks, Evaluability | 🗸 | 🗸 |  |  | 🗸 |
| Australian government, Australia | 🗸 | 🗸 | 🗸 |  | 🗸 |

\*Primary and secondary search terms were used in every search, with tertiary terms separately added as additional terms thereafter

1. The inclusion and exclusion criteria outlined in Table 2 were applied to search results.

**Table 2. Review inclusion and exclusion criteria**

| **Inclusion criteria** | **Exclusion criteria** |
| --- | --- |
| * Systematic reviews OR primary research/program evaluations for: * Australian veterans * Publicly funded or administered health programs (both state and federal) * Health conditions relevant for veterans * Australian data, studies and reports. | * Published prior to 2015\* |

\*This time period was extended for the search beyond the ‘typical’ 5-year period to account for the impact of the COVID-19 pandemic, which has skewed the research landscape over the last 2 – 3 years.

1. The title and abstract of the results were manually filtered to identify health programs that were similar to DVA health programs, recognising there were no directly equivalent programs in the Australian landscape. Health programs were considered similar if they addressed similar conditions, services or populations served compared with existing DVA programs (as shown in Table 3) and were included for review. For example, titles and abstracts indicating that the program evaluated related to medicines use or PBS expenditure were considered relevant for DVA Program 2.3 – Veterans’ Pharmaceutical Benefits, and was included for review.

Table 3. Characteristics of health programs selected for inclusion in the literature review, mapped to existing DVA programs

|  |  |
| --- | --- |
| **Health program characteristics of literature included in this review e.g. conditions, services or populations served** | **Relevant existing DVA program**\* |
| Medical services including GP and specialist services | Program 2.1 – General Medical Consultations & Services |
| Hospital services, including public, private and day procedures | Program 2.2 – Veterans’ Hospital services |
| Medicines use or PBS utilisation | Program 2.3 – Veterans’ Pharmaceutical Benefits |
| Community care and support including nursing | Program 2.4 – Veterans’ Community care and support |
| Veterans counselling & other services | Program 2.5 – Veterans’ Counselling & other Health Services |
| Rehabilitation services | Program 2.6 – Military Rehabilitation & Compensation Acts – Health & Other Care Services |
| Disability support | Program 1.2 - Veterans’ Disability Support |
| Prevention & Early intervention services including health checks | Mental and Social Health Programs – Staying Well, Prevention & Early intervention |
| For populations at increased risk, example conditions include heart health, PTSD, mental health | Mental and Social Health Programs – For those at increased risk |
| For populations with mild - moderate needs example conditions include mental health, alcohol & other substance disorders, mefloquine exposure | Mental and Social Health Programs – For those with mild-moderate needs |
| For populations with complex and severe needs e.g. trauma, PTSD, homelessness | Mental and Social Health Programs – For those with complex – severe needs |

\*Based on additional background materials provided by DVA(1)

1. The snowball technique, where reference lists of included literature are reviewed, was then used to identify any additional relevant literature.
   * 1. Grey literature search

The following outlines the protocol for the grey literature search:

1. Google was used to search for publicly available evaluation reports and guidelines on jurisdiction websites and webpages of well-regarded evaluation sites, as per Table 4. These sites were reviewed for:

* Policies or guidelines to inform evaluation
* Evaluation approaches used
* Evaluation reports for programs that are applicable or specific to DVA programs or veterans’ healthcare needs using the same criteria as the published literature search.

1. Like the Published Literature search, information before 2015 was excluded.

**Table 4. Organisations reviewed in the grey literature search**

|  |  |  |
| --- | --- | --- |
| **Jurisdictions** | **Evaluation Societies** | **Peak Bodies / Experts in the field** |
| * Australian Government Department of Health * Australian Department of Finance (now Australian Centre for Evaluation) * NSW Ministry of Health * QLD Department of Health * Victorian Department of Health and Human Services | * Australian Evaluation Society * UK Evaluation Society * American Evaluation Association | * Better Evaluation * Centres for Disease Control and Prevention (US) * The Kings Fund (UK) |

* 1. Key informant interviews

Key informant interviews were used to identify and understand the transferability of learnings from the US, which is mature in conducting evaluations for veterans’ programs. The US was identified as a suitable jurisdiction given the large number of health services and programs available to veterans. Leveraging Australian EY team’s international contacts, the team identified potential interviewees and confirmed their suitability with DVA. Once interviewees were confirmed, the team:

1. Contacted interviewees to introduce them to the project and invited them to participate in the interviews
2. Developed discussion guides to address the evaluation questions
3. Scheduled and conducted 1-hour interviews over MS Teams (note that DVA project team members were also invited to attend). The team also took detailed notes during the interviews for thematic analysis
4. Developed interview summaries and conducted thematic analysis (see Appendix B- Key informant interviews )
   1. Case studies

Case studies were used to identify key learnings from Australian organisations identified as being experienced in conducting health program evaluations, to provide in-depth insights into:

* Evaluation practices, including changes to methodology over time and operational aspects such as resourcing, workflow planning.
* Organisational considerations for conducting evaluations such as office setup, resourcing, capability building, challenges encountered and keys to success.
* Applicable lessons learned and what this means for DVA.
* The cost of conducting evaluations, including the role of conducting them using independent agencies versus internal resources/teams.

To carry out the case studies, the team:

1. Identified suitable candidate organisations using a combination of results from the desktop literature review as well as EY and DVA staff experience and connections. Selected organisations were approached and invited to participate.
2. Prepared a discussion guide and shared with participants beforehand.
3. Scheduled interviews with nominated representatives from willing organisations and conducted 30-minute interviews over Microsoft Teams. Notes were taken during the interviews for thematic analysis.
4. Prepared one-page summaries for each organisation and undertook thematic analysis (see Appendix C for case study summaries).
   1. Review search results

Table 5 outlines the results of the desktop literature review, and details of the key information interviews and case studies. Table 6 maps the document type and author to the references listed in Section 5 (References). Figure 2 indicates how each data source informed each review question. Note that not all articles identified in the desktop literature review could contribute to answering the review questions.

Table 5. Summary of data sources generated from the review approach

|  |  |  |
| --- | --- | --- |
| **Desktop literature review** | **Key informant interviews** | **Case studies** |
| * Over 500 articles generated from initial the search strategy across published and grey literature * Application of the manual filter resulted in a total of 45 articles: * 17 evaluation and policy documents\*(Appendix A: Table 16) * 16 articles that published the results of evaluations on health programs relevant to DVA health programs  (Appendix A: Table 17) * 12 systematic reviews of evaluations of health programs relevant to DVA health programs  (Appendix A: Table 18) | 2 key informant interviews conducted with US experts:   * EY Executive Director of Technology Consulting, Data & Analytics (United States) (Appendix B:Table 19) * EY Partner of Strategy and Transactions, former Assistant Secretary for Enterprise Integration for Department of Veterans Affairs (United States) (Appendix B: Table 20) | Representatives from three Australian organisations interviewed   * Program evaluation unit, NT Department of Treasury and Finance (Appendix C: Table 21) * NSW Ministry of Health Advisor (Appendix C: Table 22) * Australian Government Department of Finance (now Australian Centre for Evaluation) (Appendix C: * Table 23) |

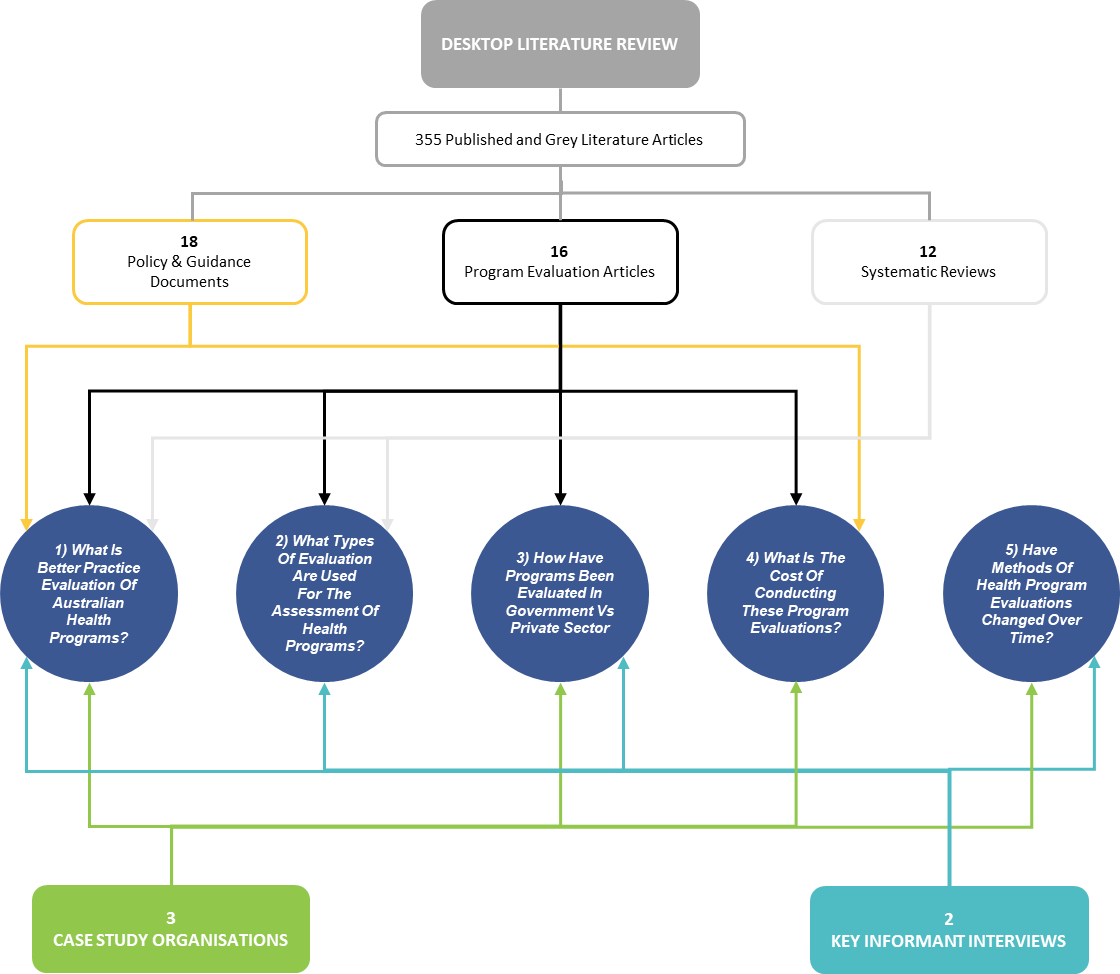
\*We reviewed all websites listed in Table 4, however, not all sites provided program guidelines and policies to inform evaluation. Therefore, only those that provided specific guidance were used and referenced in this report.

Table 6: Document type and author of references in Section 5 - References

| **Policy and Guidance documents** | | **Program evaluation articles** | | **Systematic reviews** | |
| --- | --- | --- | --- | --- | --- |
| Ref no.\* | Author | Ref no. | Author | Ref no. | Author |
| (2) | Australian Government Department of Finance (now Australian Centre for Evaluation) | (3) | Barton, Christian J. et al. | (4) | Barclay, Linda et al. |
| (5) | Australian Government Department of Finance (now Australian Centre for Evaluation) | (6) | Blignault, Ilse et al. | (7) | Browne, Jennifer, et al. |
| (8) | Australian Government Department of Finance (now Australian Centre for Evaluation) | (9) | Campbell, Brayden et al. | (10) | Dabkowski, Elissa et al. |
| (11) | Australian Government Department of Veterans' Affairs | (12) | Doran, Christopher M. et al. | (13) | Davis, Jenny et al. |
| (1) | Australian Government Department of Veterans' Affairs | (14) | KPMG | (15) | Deans, C |
| (16) | Australian Government Productivity Commission | (17) | Marsden, Elisabeth et al. | (18) | Geia, Lynore et al. |
| (19) | Centers for Disease Control and Prevention | (20) | Moreton, Sam et al. | (21) | Lee, Xing Ju et al. |
| (22) | Centers for Disease Control and Prevention | (23) | Perceval, Meg et al. | (24) | Lim, Megan S. C. et al. |
| (25) | Edwards, Barry et al. | (26) | Rahja, Miia et al. | (27) | Lokuge, Kamalini et al. |
| (28) | NSW Government Department of Premier and Cabinet | (29) | Scott, Theresa et al. | (30) | Newton, Danielle et al. |
| (31) | NSW Ministry of Health, Centre for Epidemiology and Evidence | (32) | The George Institute of Global Health, EY, University of Sydney | (33) | Senanayake, Sameera et al. |
| (34) | NSW Ministry of Health, Centre for Epidemiology and Evidence | (35) | Tjalaminu, Mia et al. | (36) | Snijder, Mieke et al. |
| (37) | Queensland Government Statistician's Office, Queensland Treasury | (38) | Trankle, Steven et al. |
| (39) | Social Policy and Research Unit (Superu) | (40) | University of Queensland |
| (41) | UK Evaluation Society | (42) | University of Melbourne |
| (43) | Victorian Government Department of Health, Prevention and Population Health Branch | (44) | University of Queensland |
| (45) | Volkov, Boris B. et al. |
| (46) | Australian Government Department of Finance (now Australian Centre for Evaluation) |

\*As per references list, Section 5

Figure 2. Data collected mapped to review questions



1. Key findings

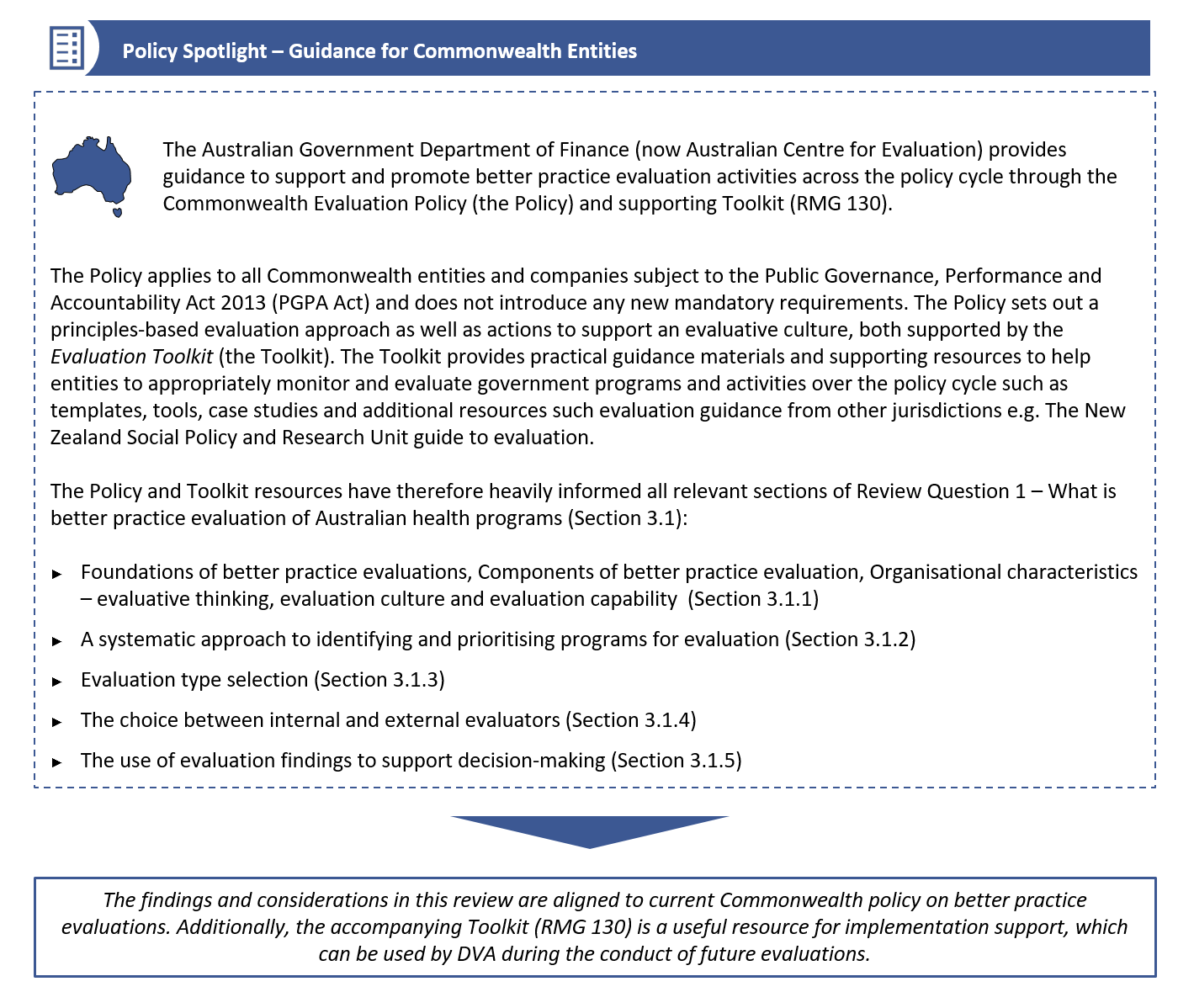
This chapter presents key findings of the review mapped against the five review questions. Findings reference analysis outputs, which are provided in the Appendices. The findings are largely informed by Australian-based literature and case studies and identifies any key lessons that may be learned from international literature and key informant interviews.

* 1. What is better practice evaluation of Australian health programs?

This section was informed by the policy and guidance documents, the program evaluation articles and systematic reviews that discussed better practice evaluation practices. Key informant interviews with experts in evaluations for veterans in the United States and case studies with Australian government organisations also informed this section. Five principles that characterised best practice evaluations were identified and are outlined below.

1. The right foundations are used
2. A systematic approach is used when prioritising programs for evaluation
3. Evaluation type is selected based on need or intended goals
4. Careful consideration is given when choosing between internal and external evaluators
5. Evaluation findings are used to support decision making.

A key resource that informed these principles was the ‘Commonwealth Evaluation Policy’ (46), which is described in more detail in Figure 3.

Figure 3. Policy spotlight - Better practice evaluation guidance for the Commonwealth

* + 1. The right foundations are used

The review identified that, separate to the implementation of an evaluation, several environmental and organisational aspects act to support and enable successful evaluation of health programs. These are described below.

Key principles of conducting high quality evaluations

This section was informed by policy documents including those from Queensland Treasury (37), Australian Government Department of Finance (now Australian Centre for Evaluation) (5) (2) (8), the New South Wales Ministry of Health and NSW Government Department of Premier and Cabinet (28). Evaluations conducted on health programs in Australia are largely underpinned by a number of principles to support leading practice. These include:

1. ***Embedding evaluation into program design*** to improve the quality of the evaluation by clearly defining measurable outcomes, considering data collection strategies and planning for organisational requirements, capacity building and methodology planning (28) (37).
2. ***Basing evaluation on sound methodology*** and ensuring that evaluations are fit for purpose, recognising a program’s size, strategic significance, risk, value and impact (28) (6) (37).
3. ***Including adequate resources and time to evaluate***, considering budget and timings that support successful evaluation e.g. realistic timeframes to allow for approval processes and reporting requirements (28) (3).
4. ***Using the right mix of expertise and independence***, including ensuring that evaluations involve credible, experienced, technically and culturally capable evaluators who are independent from program managers. This then ensures that all evaluation activities exercise sensitivity and respect towards different beliefs, perspectives and cultures (28) (6).
5. ***Ensuring proper governance and oversight*** by using adequate governance and risk management processes to ensure oversight and risk management of evaluation design, implementation and reporting. Governance structures should be built into the evaluation strategy (28).
6. ***Being ethical and culturally appropriate in designing and conducting the evaluation*** and ensuring there is no physical, psychological or reputational harm arises as a result of the evaluation. This can be done by protecting participant privacy, and upholding ethical standards of conduct, for example carefully considering ethical implications of any evaluation activities particularly around data collection or personal data and especially when assessing for impacts on vulnerable groups. This recognises that evaluations often present a number of moral, ethical and political concerns. This may require formal review and approval from an ethics committee certified by the National Health and Medical Research Council (28) (6) (3).
7. ***Being informed and guided by relevant stakeholders*** including program participants, government or non-government staff or agencies involved in managing or delivering the program, or senior decision makers (28).
8. ***Using evaluation data meaningfully***, by ensuring results are clearly articulated in evaluation reports, and in turn using reports to inform decision making (28) (6).
9. ***Being transparent and open to scrutiny*** by publicly releasing key information about all aspects of the evaluation (such as methodologies, assumptions, analyses and findings) as a default, unless there is overriding public interest against disclosure. This includes release to appropriate stakeholders (28) (6).

These key principles were largely supported by the Department of Finance (now Australian Centre for Evaluation), who emphasised the importance of conducting evaluations that are robust and ethical, using evaluation findings meaningfully to guide decision making, planning evaluations appropriately to make sure evaluation activities are proportional and appropriate and understanding the risk landscape of programs to help prioritise evaluations (Table 23: Case Study III). The NSW Ministry of Health strategic advisor also emphasised the benefits of considering a high-value healthcare framework when evaluating health programs; there are several key components of quality healthcare, which include considering the different stakeholder perspectives, effects on longer term health outcomes and importance of striving towards equitable healthcare e.g. culturally appropriate care (Table 22: Case Study II).

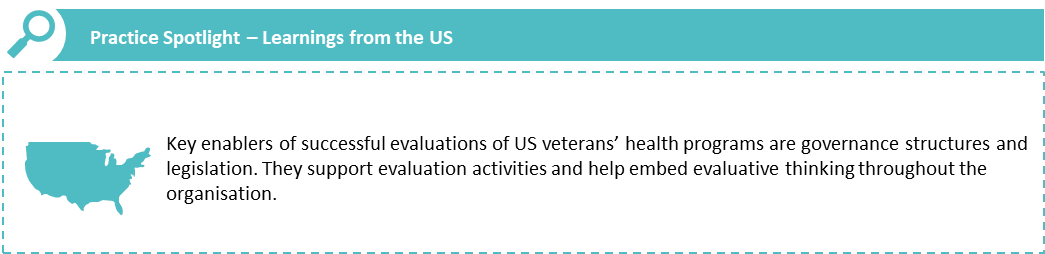
Figure 4 outlines the principles of better practice evaluations from the UK Evaluation Society (41), and Figure 5 outlines a key enabler in successful veterans’ programs evaluations, as identified by US expert as part of the Key informant interviews (Appendix B)

Figure 4. Learnings from the international literature – Principles for sound evaluation (41)



Data collection source: Desktop review (41)

Figure 5. Learnings from the US – A key enabler for successful evaluation of veterans’ health programs

  
Data collection source: Key Informant Interviews

Components of a better practice evaluation

A range of policy and guidance articles and select program evaluation articles noted three key components of better practice evaluation, which are outlined in Table 7. These are logic models, evaluation frameworks and evaluation plans.

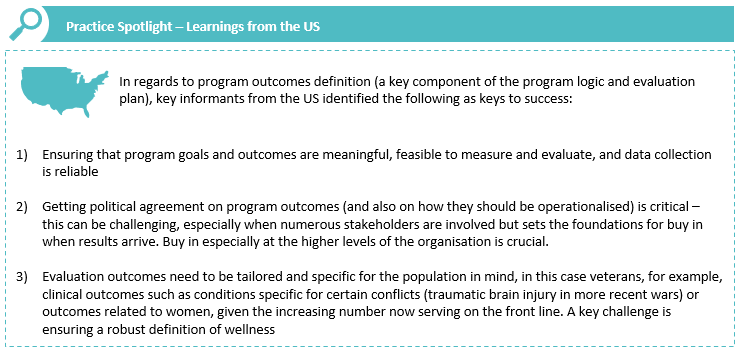
Table 7. Supporting components of better practice evaluations (34) (28) (3) (6) (38) (43) (5)

| Components | How it supports better practice evaluations |
| --- | --- |
| **Logic model / program logic\*** | A management tool that presents the logic of a program in a diagram or schematic (with related descriptions). It illustrates how a program is intended to work, showing the logical linkage between the identified need or issue that a program is seeking to address; its intended activities and processes; their outputs; and the intended program outcomes. The explicit representation of the assumed logic creates an opportunity for stakeholders to understand and test the assumptions, ensuring that the program design is sound.  Program logics allow for a better understanding of the broader drivers of a program’s efficacy, they can focus the research team on the intended outcomes, and can serve as a solid foundation throughout the program implementation and evaluation. They can be used to scrutinise use and allocation of resources and help tailor programs for local context. Additionally, the different segments of the program logic can be allocated indicators to measure progress of a health program. |
| **Evaluation framework** | A tool that links what the program intends to achieve, and how success can be measured (that is, what are the indicators that need to be measured to determine whether change has occurred). |
| **Evaluation plan** | An evaluation plan outlines the objectives of the evaluation and clarifies the roles and responsibilities of those involved. It sets out the most appropriate evaluation design throughout the project, and the appropriate data collection methods. It also clarifies assumptions on which the program design and implementation were based. The evaluation plan should also include how the evaluation results will be disseminated and the cost of the evaluation. It should be developed for all new programs before they are implemented, and it should be written alongside the overall program plan. |

The case study consultation with the NSW Ministry of Health indicated that program logics are most useful for programs with outcome timelines (for example, have both short and long-term outcomes). They are also able to pick out faults with the program, for example, design problems with the program. Program logics can also be used to clarify assumptions of the program and can help guide the identification of desired outcomes over time (Table 22: Case Study II).

Developing program logics and evaluation frameworks requires well-defined program outcomes. Figure 6 outlines important steps to successfully defining intended program outcomes according to a key informant interviewee. These lessons can be transferred to the Australian context.

Figure 6. Learnings from the US - successfully defining program outcomes



Data collection source: Key Informant Interviews

Organisational characteristics – evaluative thinking, evaluation culture and evaluation capability

Successful evaluation requires more than the technical skills and knowledge. To harness the full impact of an evaluation, organisations require an evaluative culture with corresponding evaluation capability (37) (25), which is supported and enabled by evaluative thinking (5).

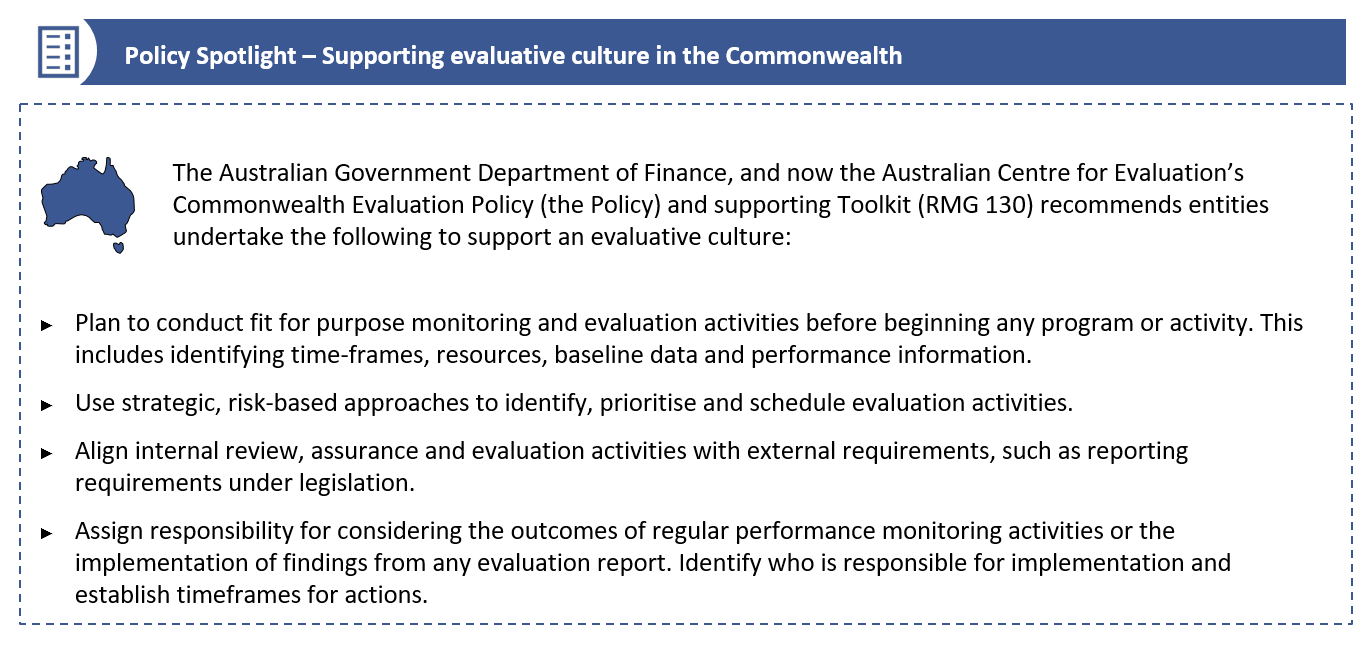
Evaluative thinking is defined as a *form of critical thinking where evaluative questions are asked as a part of everyday business* (5) or in some cases, *a constant state-of-mind within an organisation’s culture and systems to reflect, question, learn and modify using the evidence available* (19). Practically, this involves embedding evaluative thinking into the way people work and think (5) as well as all parts of the program’s lifecycle (37), and can also involve the use of robust analytical methods to continuously assess program performance against stated objectives (5). It is therefore logical that evaluative thinking in turn supports an overarching culture of evaluation in an organisation, where staff are encouraged to self-reflect on their activities, look for better ways of doing things, share knowledge and learn from both good practice and mistakes (5).

A strong, evaluative culture, which promotes evaluative thinking within an organisation is key to supporting ongoing improvement in evaluation processes and organisational capability (37) (5). There are several activities that organisations can undertake to build a culture of evaluation. These include:

* Having an evaluation capacity building plan and adapting capability building processes to suit their nature, size and structure
* Allocating adequate resources to evaluation skills and knowledge development
* Ensuring their program, group, office and departmental leadership teams play an important role in nurturing and fostering this culture, through explicit and ongoing support
* Establishing peer learning structures
* Accessing quality evaluation information resources
* Obtaining buy-in from all staff
* Establishing required infrastructure to support evaluation processes and communication, including necessary governance structures (37) (25) (45).

Figure 7 outlines additional activities recommended in the Commonwealth Evaluation Policy (46) to support the development of organisational evaluative culture.

Figure 7. Policy spotlight - evaluative culture in the Commonwealth (5)



It is intuitive to draw the link between evaluation capability, specifically the ability of staff to understand and undertake all key elements of the evaluation process (from designing to implementing and reporting) and the production of high-quality evaluations. However it is important to recognise that the establishment and development of evaluation capability is dependent on a strong organisational culture which values evaluative thinking, and embeds evaluation activities into organisational practice with the right policies, procedures, people and equipment to support (37) as well as invests both time and resources in training and development (37) (25).

* + 1. A systematic approach is used when prioritising programs for evaluation

This section was informed by policy documents by Australian Government Department of Finance (now Australian Centre for Evaluation) (2) (5) (8) and NSW Government Department of Premier and Cabinet (28). It is not feasible, appropriate or cost-effective to evaluate all health programs, recognising that in some cases performance monitoring alone will meet Commonwealth performance reporting requirements (8). Therefore, it is recommended that program managers take a strategic, risk-based approach to identifying and prioritising programs for evaluation, based on program size, strategic significance and risk profile (28) (8).

There are no definitive criteria for identifying programs appropriate for evaluation. However, several program characteristics can be considered when deciding whether to evaluate a program. These are detailed in Table 8.

Table 8. Program characteristics that can inform the decision on whether to evaluate a health program (28) (8)

| **Program characteristic** | **Details** |
| --- | --- |
| Priority | * Is the program a government initiative or directly a strategic priority or connected to a government priority? * Is there a Cabinet or ministerial directive to undertake a comprehensive evaluation? * Is there a Ministerial Statement of Expectations? * Is the program a priority issue for the sector or other key stakeholders? * Does the program have an important relationship to other program areas? |
| Program accountability | * At what level of government is ultimate accountability for this program? |
| Funding | * What is the degree of government or departmental investment in this program? |
| Scope | * How many entities are involved in this program, including government departments, agencies or external delivery partners? |
| Risk | * How high is the program’s risk e.g. budget, operations? * What are the risks of conducting or not conducting the evaluation? |
| Monitoring, review and stakeholder feedback | * Have issues with the activity or program objectives, implementation or outcomes been identified through monitoring, review or stakeholder feedback? * Does performance information suggest areas where improvements are required? * Has termination, expansion, extension or change in the activity or program been proposed? |
| Profile and sensitivity | * Does the activity, program or evaluation topic have a high profile, high sensitivity or a high cost? * Has it attracted significant public attention or criticism? |
| Previous evaluations | * Has the program been evaluated previously?   This may help determine if a new evaluation is worthwhile, particularly if:   * significant amount of time has lapsed since the previous evaluation, * the previous evaluation pointed to a need for change, or * there has been a significant change in the program or activity. |
| Other factors | * A lack of evidence base * Reporting or evaluation requirements including regulatory or legislative requirements, evaluation commitment in a new policy proposal, an Australian government budget or public statement |

* + 1. Evaluation type is selected based on need or intended goals

Policy and guidance documents identified that three main types of evaluation are commonly conducted for health programs in Australia: Process, Outcome and Economic evaluations (Table 9). It is noted that this is not an exhaustive list, rather the results of the review search strategy.

Table 9. Evaluations commonly used in Australian health programs (28) (22) (39) (43)

|  |  |  |
| --- | --- | --- |
| **Evaluation type** | **Purpose** | **Program aspects examined** |
| **Process** | To determine if the program activities have been implemented as intended | Operations, processes and service delivery |
| **Outcome** | To identify the effectiveness of a program in achieving desired outcomes in the target populations. This can be done in both the short and long term | Progress and achievement of outcomes, who the program works best for and the ideal circumstances. |
| **Economic** | To identify, measure and value a program’s economic costs and benefits | Reliable results data for both program inputs and outcomes |

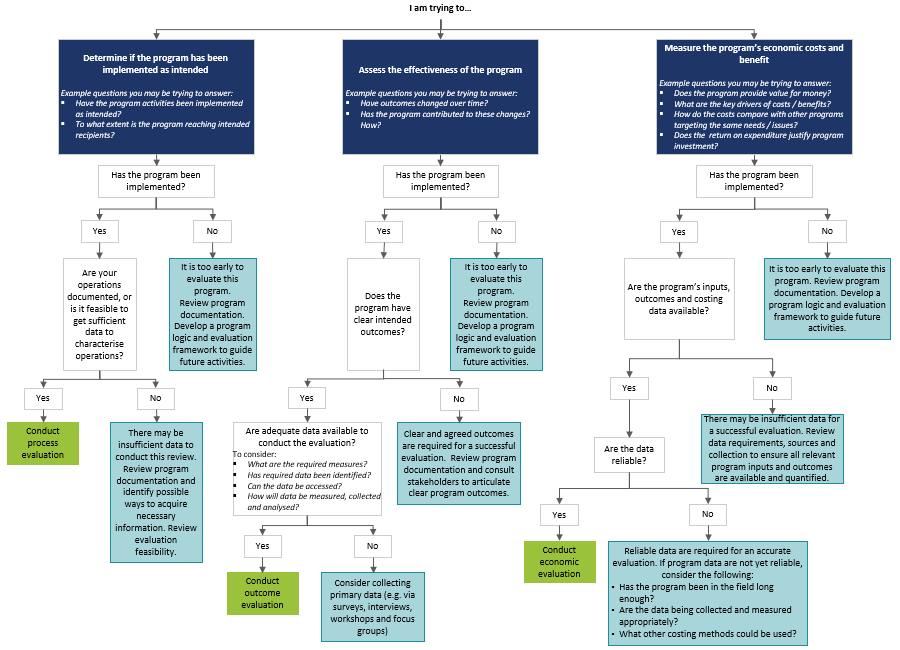
Choosing which of these evaluation types to use is largely informed by the evaluator’s purpose for the evaluation, the type of information required from the evaluation and the intended use of the results. These, and other considerations for choosing evaluation type are outlined in Table 10.

Table 10. Evaluation type characteristics (28) (22) (39) (37) (43)

|  | **Process** | **Outcome** | **Economic** |
| --- | --- | --- | --- |
| **When to use it with respect to the program lifecycle** | * Can be done as soon as the implementation begins or during operation of an existing program | * After the program has made contact with at least one person or group in the target population * Ideally when the program has been running long enough to produce reliable results | * At the beginning of a program * During the operation of an existing program |
| **What it shows** | * How well the program is working, the extent to which it is being implemented as intended * Whether the program is accessible or acceptable to the target population | * The degree to which the program is having an effect on the target population’s behaviour i.e. the effectiveness of the program * Can identify who the program works best for and under which circumstances | * What resources are being used, their direct and indirect costs compared with outcomes |
| **Uses** | * To provide an assessment of program activity performance * To track reach and level of implementation * To inform adjustments to service delivery and program redesign * To identify potential or emerging problems | * Shows whether a program is effective in meeting its outcomes-based objectives | * To promote efficient and appropriate resource allocation * To compare alternatives with regard to public or economic value e.g. no policy change option, compared to other programs * To inform whether a program should continue or cease |
| **Methodologies used** | * Business and process analysis techniques including: * Document reviews * Observation * Surveys * Individual or group interviews * Administrative program data analysis | * Experimental, quasi-experimental and non-experimental designs | * Cost benefit analysis, cost effectiveness analysis, cost analysis, cost utility analysis, cost minimization analysis, multi-criteria analysis, cost consequence analysis, return on investment, economic impact analysis |
| **Requirements to conduct** | * Program has been operating long enough to document and assess operations | * Program design complete i.e. outcomes specified, measurement tools in place and used change is clearly set out with Program logic) * Program data identified, collection and management organised, data analysis methodology confirmed, how the data will be reported (this includes any comparison data) * Organisational and cultural context | * Specialist evaluators to conduct the following * Can only be done when a program is producing reliable results data that can be valued (needs planning before implementation) |

Figure 8 on the next page provides a decision-tree for choosing between a process, outcome and economic evaluation.

Figure 8: Decision tree for choosing between a process, outcome and economic evaluation



The NT Department of Treasury and Finance guides Northern Territory agencies in choosing the most appropriate type of evaluations dependent on the goal of the evaluation. They acknowledge that, while all evaluation types can offer value to organisations, process evaluations are the most commonly conducted type of evaluation, as many Northern Territory agencies are still at the beginning of their evaluation journey (Table 21: Case Study I).

Outcome evaluations require some special consideration, as they tend to use (often) sophisticated quantitative methods. Guidance from the NSW Ministry of Health and selected articles from the published literature review identified three types of study designs can be employed for outcome evaluation, which vary according to the stage of implementation where they can be used – randomised controlled trials (RCTs), quasi-experimental designs and non-experimental designs. Table 11 outlines the features of different outcome evaluation designs.

Table 11. Comparison of study designs used in outcome evaluation (34) (24) (27)

|  |  |  |  |
| --- | --- | --- | --- |
| **Study design** | **Randomised Controlled Trials (experimental design)** | **Quasi-experimental designs** | **Non-experimental designs** |
| **Description** | The establishment of two or more groups that are identical except for a single factor of interest, for example, exposure to an intervention. Individuals are randomly assigned to either the control group or an intervention group at the start of the study to measure the impact of the intervention. | Controlled before and after studies measure observations before and after the implementation of an intervention and compare them to a control group who did not receive the intervention. The main difference between this design and an RCT is that there is no random assignment to treatment groups. A matched control group (consisting of non-participants) is used for comparison purposes and helps to ascribe the effect to the intervention itself. | Unlike experimental designs, non-experimental designs do not include a control group whereby participants or groups are randomly assigned to receive the intervention or usual care, nor is there randomisation in terms of sampling. |
| **When to use this method** | Prior to program implementation. This study design is not plausible if the intervention has already started, unless testing an enhancement of the program | Prior to program implementation. This study design is not plausible if the intervention has already started, unless testing an enhancement of the program | At any point in the program lifecycle. This method should be used only after other design possibilities have been considered. |
| **Strengths** | * The most robust way to evaluate the effectiveness of a new treatment or prevention strategy * Best study design when trying to establish causality, hence they are most relevant when there is a need to generate causal evidence | * Relatively simple and cost-effective to undertake (may provide a compromise where RCTs are not possible) | * Can evaluate interventions where there is no control group * Can enable the long-term effects of population-level interventions and policies in a real-world setting to be evaluated * Can allow for both prospective and retrospective cohort studies |
| **Specific limitations** | * Costly to design and deliver * An adequate sample size is required to ensure that a difference between the two groups is detected * Often not possible to randomise cohorts * May involve ethical risks, such as withholding a program intended to improve health outcomes from one group | * Difficult to determine, or control for, the impact of other influences or events on the outcome * Can be time consuming and expensive | * Difficult to establish cause-and-effect * Typically require a large sample size and long follow-up period in order to determine the associations between exposures and outcomes * Retrospective studies rely on past data collection, and as such the data may be incomplete, inaccurate or inconsistently measured between subjects * Prospective studies are time-consuming and expensive |

The NSW Ministry of Health outlined that it is important to have a baseline to compare endpoint outcomes against. It is important for the evaluator to separate the effects of the program from natural progression in order to articulate the benefits of the program (Table 21: Case Study I). However, doing this via RCTs is often difficult, as acknowledged by the NT Department of Treasury and Finance, who currently do not have the capability to undertake RCTs at this stage of their evaluation journey (Table 21: Case Study I).

One of the benefits of quantitative data is the ability to use the data to perform data analytics. This was noted by DVA in the case study consultation with the NSW Ministry of Health, they have a closed system of providers and consumers. This, alongside high quality, longitudinal utilisation, and costing data, places them in a strong position for new data analytics, for example, dynamic simulation modelling or forecasting for burden of ill health (Table 22: Case Study II).

Guidance from the Queensland Treasury and selected articles from the published literature review indicated that a mixed methods approach to data collection, using both quantitative and qualitative data is ideal (37) (7) (24) (27). This was also supported through case study consultations with the NSW Ministry of Health, who noted that a ‘single number’ approach to collecting data is phasing out, and mixed methods is an ideal approach to data collection (Table 22: Case Study II). This mixed-methods approach allows for diversity in feedback and can better identify program areas which are doing well, need improvement, and also how these areas can improved (7) (24) (27).

Mixed method approaches can measure the impact of health programs and can contextualise cultural inferences about a targeted health program. Strong collaboration between quantitative and qualitative evaluators is necessary for a strong and cohesive evaluation, particularly for experimental evaluation designs. For any type of evaluation, valid reliable tools that can accurately measure output are critical, and appropriate comparison where possible to non-intervention settings is vital for any high-quality evaluation to occur (18) (30).

While quantitative and experimental evaluation designs are highly regarded, qualitative data enables a participative and collaborative approach to evaluation and is incredibly valuable to contextualise and provide culturally relevant inferences about a program’s utility and impact (27). Involving key stakeholders and participants in the designing of the program and its evaluation process can help improve the data and outcomes of a program. Qualitative data are most commonly collected via interviews (mainly semi-structured) and focus groups, and analysed via thematic analysis (10) (29) (33) (36) (35) (38). Of note is that culturally appropriate language is fundamental when engaging with vulnerable populations for qualitative data collection (6).

There are several unique qualitative evaluation techniques that can be used, including the Most Significant Change (MSC) technique, used in an *Evaluation of the National Empowerment Project Cultural, Social, and Emotional Wellbeing Program* (35).It is a robust methodology to collect and analyse qualitative data, enabling evaluators to determine and measure intangible qualitative indicators of importance. The technique is valuable in measuring outcomes of social change programs and is particularly relevant when those social change programs hold seemingly ‘intangible’ goals, such as increasing confidence. The MSC technique is suitable for researchers who:

* have regular contact with participants,
* wish to collect data on impact rather than outputs, or
* want to investigate any possible unintended consequences to their program (35).

Guidance from the Queensland Treasury and selected articles from the published literature review indicated that a mixed methods approach to data collection, using both quantitative and qualitative data, is ideal (37) (7) (24) (27). This was also supported through case study consultations with the NSW Ministry of Health (Table 22: Case Study II). A mixed-methods approach allows for diversity in feedback and can better identify program areas which are doing well or need improvement, and also how these areas can improved (7) (24) (27). Mixed method approaches can measure the impact of health programs and can contextualise cultural inferences about a targeted health program.

Strong collaboration between quantitative and qualitative evaluators is necessary for a strong and cohesive evaluation, particularly for experimental evaluation designs. For any type of evaluation, valid reliable tools that can accurately measure output are critical, and appropriate comparison where possible to non-intervention settings is vital for any high-quality evaluation to occur (18) (30).

When deciding the use either qualitative or quantitative data collection methods, DVA could consider using both, using Table 12 as a guide.

Table 12. Rationale for quantitative and qualitative data collection (34) (36) (27)

|  |  |
| --- | --- |
| **Use quantitative data collection methods when:** | **Use qualitative data collection methods when:** |
| * Differences or change in an impact or outcome needs to be quantified * Validated and reliable measures are needed to answer an evaluation question * Causal evidence of program effects is needed (it is important to note that the program is rarely the sole cause of change; there may be other activities or environmental factors which provide partial attribution) * Data are needed on a large number of people or populations | * Quantitative findings need to be strengthened with a deeper understanding of unique contexts * Need to analyse and explore complexity, meaning, relationships and patterns * A deeper understanding of cultural facilitators and barriers to a health program is needed |

* + 1. Careful consideration is given when choosing between internal and external evaluators

Guidance from the Australian Government Department of Finance (now Australian Centre for Evaluation) (2) (5) (8), and the NSW Ministry of Health and Queensland Treasury (37) was used to inform this section. Determining who should manage an evaluation will depend on a variety of factors, including resourcing requirements, capabilities, the need for independence and strategic significance, degree of risk and the complexity of the program (37) (31). Organisations have the choice to either conduct the evaluation in-house, or commission an external evaluator to support the evaluation process. Ideally, to ensure that program and evaluation planning can occur simultaneously, an evaluator should be chosen during program development to support development of an evaluation framework (37).

Considerations when deciding between conducting an evaluation in-house (with an internal evaluator) or commissioning an external evaluator are outlined below in Table 13. In summary, engaging an external evaluator should be considered where there is a need for special evaluation expertise, and/or an independent assessment of the program. An external evaluator is likely to be important for programs that have involved a reasonable investment, where independence is more desirable; also for programs being assessed for continuation, modification, or scaling up (31). The use of internal evaluators has the potential to be less costly and can help build internal evaluation capability and an evaluation culture within an organisation. However, internal evaluations can draw resources away from program delivery and can be biased in favour of the program. The NT Department of Treasury and Finance uses both in-house and external evaluators, however, most of the program evaluations conducted are commissioned to external evaluators as they are still building their evaluation capability (Table 21: Case Study I).

Table 13. Comparison between internal and external evaluators (37) (31) (39)

|  | **Internal Evaluator** | **External Evaluator** |
| --- | --- | --- |
| **Advantages** | * Has detailed knowledge about program design and implementation * Can help to build evaluation capacity, knowledge and skills within the organisation * Can be less costly * May be able to achieve better employee and stakeholder ‘buy-in’ for the evaluation | * Ability to provide independence and objectivity (which is needed for programs that are subject to higher scrutiny and risk) * Offers specialist technical or professional skills and expertise * Can evaluate more complex programs * Provides their reputation and experience * May deliver more efficient evaluation activities * Lesser impact internal operational functions / resources than an internal evaluation |
| **Disadvantages** | * May have less evaluation experience * May draw resources away from program delivery, or other internal activities * May inhibit the candour of stakeholders who are consulted as part of the evaluation * May reduce the actual or perceived validity of the evaluation’s results | * Can be more costly * May not have an adequate understanding of the program, its implementation context or target groups * May have difficulty engaging employees and stakeholders in the evaluation process |

* + 1. Evaluation findings are used to support decision-making

Evidence-informed decision-making should be at the forefront of any program evaluation (37), and program evaluation should be designed to be used constructively to support continuous improvement (2). Evaluation findings can help support decision-making by policy makers about a program’s design, implementation and opportunities for improvement. It can also aid in the understanding of the information needs of other evaluation stakeholder groups, including those who may be affected by its findings (37).

Government evaluation guidelines indicate that organisations are encouraged to share the results of evaluations in the public domain (where permission has been obtained). Sharing results publicly enables the rich, evidence-based insights gathered through the evaluation to help guide the allocation of public resources for government agencies, improve the design and implementation of programs, and deliver better services (37) (2). Furthermore, published evaluation findings support accountability and continuous improvement, and helps to embed a culture of evaluation (2).

The review also identified the importance of communicating changes being implemented as a result of evaluation findings, particularly with those who have participated in the evaluation. Providing feedback to participants (where disclosure is allowed) recognises and values their contribution (37). The importance of completing the ‘feedback loop’ was also highlighted through the case study consultation with the Department of Finance (now Australian Centre for Evaluation), who emphasised that this allows evaluation findings to be used in an effective way (Table 23: Case Study III)

* 1. What types of evaluations are used for the assessment of health programs across Australia?

A total of 16 Program Evaluation articles were identified during the desktop review (see Figure 2). These articles presented results of primary research/program evaluations for publicly funded/administered health programs for Australian veterans and/or people with health conditions relevant for veterans (as per Table 2). The characteristics of these evaluations is outlined in Figure 9. It can be seen that mixed methods were used by 68% of the researchers, supporting the findings presented in the previous section about mixed methods being the better practice approach to conducting evaluations. Interestingly, all of the evaluations conducted by government agencies were carried out by external evaluators, and the most common type of evaluations carried out were process evaluations.

Figure 9. Types of evaluations identified in the 16 program evaluation articles



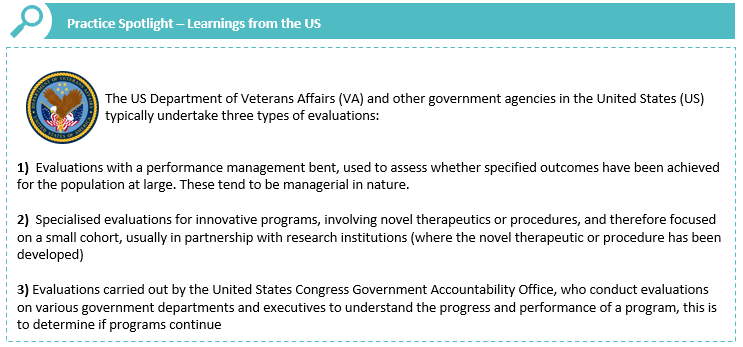
A total of 12 systematic reviews were also identified by the review. These were systematic reviews of publicly funded/administered health programs for Australian veterans and/or people with health conditions relevant for veterans. The 12 systematic reviews presented the findings of over 355 publications. In total, nine of these 12 systematic reviews provided details on the types of evaluations used in the publications included (equating to 229 publications in the systematic reviews presenting evaluation design type). Table 14 shows that the most common evaluation methods reported were pre-post design, which is usually a quantitative method that compares the outcomes of a cohort before and after an intervention is implemented. Mixed methods designs were not as common as the findings above would suggest; however this may be due to differences in reporting methods by the systematic review authors.

Table 14: Evaluation types from systematic review articles

|  |  |  |
| --- | --- | --- |
| **Evaluation Design Type** | **Number of Articles** | **Percentage (%)** |
| **Pre-post** | 89 | 39 |
| **Quantitative** | 46 | 20 |
| **Qualitative** | 35 | 15 |
| **Descriptive** | 31 | 13 |
| **Other (including pilot studies and modelling)** | 15 | 7 |
| **Mixed methods** | 13 | 6 |
| **Total** | 229 | 100% |

Figure 10 outlines some of the common types of evaluations used in the US, as identified during key informant interviews. Key is the use of specialised evaluations for innovative programs, which involve smaller cohorts in partnership with research institutions. This concept of developing evaluation methods that suit the program was supported by the NSW Ministry of Health. The NSW Ministry of Health highlighted the importance of framing data collection around the potential future use of data and making sure data collection methods were appropriate and could gather this information as directly as possible (Table 22: Case Study II).

Figure 10. Learnings from the US - veteran program evaluations in the US



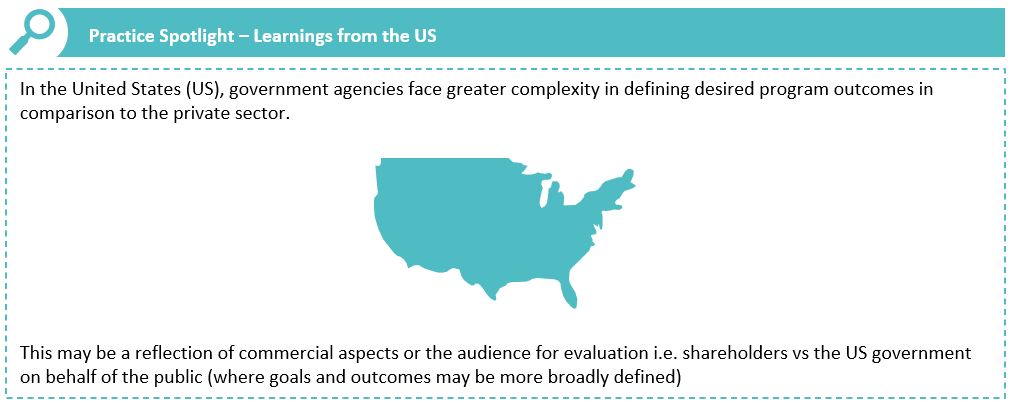
Data source: Key Informant Interviews

* 1. How have health programs been evaluated within government agencies? Does this differ from the private sector?

This section was informed by the program evaluation articles found in the published literature review and from key informant interviews. Case study discussion further supported this information. The desktop literature search identified that in Australia, all government evaluations were commissioned (9) (12) (17) (20) (23) (38). This information was reinforced by the NT Department of Treasury and Finance confirming that most of their program evaluations have been commissioned (Appendix C, Table 21). The private sector program evaluations from the published literature review had a mix of both commissioned and in-house evaluations. There appeared to be no meaningful differences in evaluation type between government and private sector agencies.

Key informant interviewees suggested that clarifying program outcomes was more complicated for government agencies than in the private sector, which may partially be due to a difference in evaluation audience (Appendix B).

Figure 11. Learnings from the US - private sector evaluations



Data source: Key Informant Interviews

* 1. What is the cost in conducting these health program evaluations?

There was very little information on the absolute cost of conducting health program evaluations in the literature – and none of the 16 program evaluation articles provided the evaluation budget relative to program costs. The review did not identify any definitive rules on how much an evaluation should cost; rather that cost should be informed by program and evaluation characteristics. This was supported through consultations with the NSW Ministry of Health and the Department of Finance (now Australian Centre for Evaluation), who acknowledged that a range of factors influence evaluations and as such, a single ‘rule of thumb’ approach may not be appropriate (Table 22: Case Study II, Table 23: Case Study III). The only resource that gave a definitive suggestion of relative evaluation costs was the NSW Government Evidence and Evaluation Guidance series, which suggested that a commissioned evaluation should cost approximately 10% of program costs (31), however this is subject to many program and evaluation factors.

Factors influencing whether an evaluation is needed, and the likely associated costs and resourcing requirements, mentioned by the Department of Finance (now Australian Centre for Evaluation) included:

* The level of risk associated with implementing and evaluating a specific program – the cost of evaluation must be weighed against the risk of not evaluating, noting in some cases robust performance monitoring by itself may be sufficient to meet the performance reporting requirements under the [*Public Governance, Performance and Accountability Act (2013*](https://www.legislation.gov.au/Series/C2013A00123)).
* A ‘rule of thumb’ approach may not be sufficient, as it fails to take account of the characteristics and risk profile of a specific program delivered in a specific context, including relevant ethical, cultural and privacy factors (e.g. where a health program delivers services to a vulnerable cohort, this would need to be factored in to the overall evaluation design).
* The timing and type of evaluation needed to support continuous improvement, accountability and decision-making needs to be determined on a case-by-case basis to ensure the overall approach is fit for purpose, having regard to relevant ethical, cultural and privacy considerations.
* Consideration of where the program is in the policy lifecycle (i.e. whether it is a new program or one that has been in operation for some time) and the extent to which new or amended monitoring and evaluation arrangements are required to generate data and robust evidence to drive continuous improvement.   
  The costs associated with conducting a specific evaluation will be influenced by the extent to which existing data and administrative systems are generating robust evidence that can be leveraged to support an evaluation.

The Department of Finance (now Australian Centre for Evaluation) highlighted that evaluations are one aspect of a broader set of assurance mechanisms used to generate robust evidence and performance insights to support continuous improvement (Table 23: Case Study III).

Table 15 provides some concrete guidance for setting evaluation budgets. It is important to recognise that the scope and level of resourcing will ultimately influence evaluation quality and usefulness (37), and decision makers must aim to balance the need to evaluate against diverting resources from the program and its clients (39).

Table 15. Program and evaluation characteristics and their impact on evaluation cost according to Queensland Treasury (37), the New Zealand Government (39) and NSW Ministry of Health (31)

| **Characteristic type** | **Characteristic** | **Impact on cost to evaluate** |
| --- | --- | --- |
| **Program** | **Scale** | * Large scale programs with significant budgets and higher complexity are generally prioritised by government agencies and will have bigger budgets for evaluation * Pilot programs may justify more spend, given the purpose is to learn for the first time what works, how, when and why |
| **Significance** | * Programs with strategic significance tend to have bigger budgets for evaluation |
| **Risk** | * Higher levels of risk confer higher levels of scrutiny and therefore can justify more investment to ensure evaluation quality and coverage. Aspects that inform program risk level include profile, funding, target audience, level of innovation or random political factors |
| **Time in field** | * Long running, established programs which have been previously evaluated may not need significant evaluation investment given the existing body of knowledge |
| **Intended use of evaluation findings** | * Where findings will be used for the following, a high level of resources may be allocated for evaluation, even for small programs: * Used to inform the decision to scale e.g. to a wider area/different client group * Generalised or used as evidence of another program’s effectiveness |
| **Evaluation** | **Internal vs External evaluators** | * Internal evaluators may appear to be more cost efficient due to the staffing budget, however this may not always be the case * External evaluators attract additional fees and time to manage to overarching process\* |

* 1. Have the methods of health program evaluations changed over time?

The Ministry of Health indicated that there has been a pivot to target wellbeing in the public sector, as support schemes are not independent of wellbeing type outcomes. This is likely driven by a change in attitudes towards mental health, and acknowledgement that personal life inevitably affects work life (Table 22: Case Study II).

The case study consultation with NSW Ministry of Health has also indicated that over the past 10 years, evaluations have shifted from being thought about as a separate external process, towards being a significant part of the continuous improvement cycle of program implementation (Table 22: Case Study II). This was validated through consultation with the Department of Finance (now Australian Centre for Evaluation), who explained the change over time from evaluations being a compliance driven ‘tick and flick’ exercise, towards a more meaningful activity with the intention of driving continuous improvement. They also outlined that moving forward, a culture change will be needed to support people to understand that effective use of evaluations can contribute to better performance management and more effective program delivery and outcomes (Table 23: Case Study III), with work underway to help embed evaluative thinking into everyday practice across the Commonwealth.

Furthermore, the Department of Finance (now Australian Centre for Evaluation) indicated that findings from the [Independent Review of the Australian Public Service](https://www.pmc.gov.au/resource-centre/government/independent-review-australian-public-service) (APS) indicate evaluation practices within the government space have declined over time. The APS Review found that at the federal level of government, few organisations have retained evaluation expertise, with external providers being used to fill skills gaps in evaluation. They are at the beginning of a long change journey to work with federal departments and rebuild evaluation capability at the Commonwealth level (Table 23: Case Study III).

Lessons learned from the US experience with evaluating veteran programs include the rise of ‘rapid cycle evaluations’, which can be undertaken while programs are being implemented (Figure 12). These are sometimes known as ‘developmental evaluations’, which can provide close to real-time feedback to program staff and facilitate a continuous development loop. The difference between developmental evaluations and rapid cycle evaluations are that the latter may only be a one-off occurrence – whereas developmental evaluations continue to occur throughout program implementation.

Stakeholder engagement has also increased in evaluations of US veteran programs. The importance of stakeholder engagement was highlighted by the case studies too. The NSW Ministry of Health indicated that qualitative data enables stakeholders to share their respective experiences and can help clarify stakeholder attitudes and behaviours about a program (Table 22: Case Study II). The Department of Finance (now Australian Centre for Evaluation) also noted that qualitative data enables the evaluator to tell a powerful story which makes an evaluation more influential – which is one of the main goals of an evaluation (Table 23: Case Study III).

Figure 12. Learnings from the US - the rise of rapid cycle evaluations



Data source: Key Informant Interviews

1. Considerations for better practice evaluation

This chapter provides an evidence-based, better practice approach for future health program evaluations in DVA. It was developed by drawing out the implications of findings presented in Chapter 3. The flowchart at the end of the chapter can be used to help guide DVA on the way to approach evaluations to ensure an evaluation culture.

A summary of implications of the key findings are presented below. They are cross-referenced against the relevant section in the key findings.

**Better practice evaluation of Australian health programs involves:**

* Embedding evaluation into program design with appropriate governance and oversight. Having clear program goals can help direct evaluation methodology and create meaningful and useable data. Working closely with key stakeholders can also ensure that the program is culturally appropriate and set up the health program for the highest chance of long-term success. (Section 3.1.1)
* Designing a program logic and evaluation framework are key components to a successful evaluation. They help in identifying key program components and keep focus on desired program outcomes. (Section 3.1.1)
* Creating an evaluation culture within an organisation allows for capacity building, adequate resource allocation, and general understanding of the benefits and importance of evaluation. This can allow for optimal evaluation efficiency and more broadly achieving desired program outcomes. (Section 3.1.1)
* Resources often restrict the ability to evaluate all programs; careful consideration should be taken of the potential benefits and use of evaluation in achieving broader organisational and strategic goals. (Section 3.1.2)
* Process, outcome and economic were the most common types of evaluation identified by policy and guidance documents from this review. Selection of these methods design depends on what data are available as well as the maturity of the program. For example, a process evaluation is a suitable option for a relatively new program. When choosing between quantitative and qualitative data collection methods, a mixed methods design is considered ‘ideal’. This is because it allows for more diversity in information to inform which areas of a program are performing well and which may need modification. (Section 3.1.3)
* External evaluators are useful for larger scale or high investment projects where independence is desirable. They are helpful in projects intended to be scaled up or for those requiring significant modification. Internal evaluators are generally less costly and can build internal evaluation capability and culture, however can be bring bias into evaluation results. (Section 3.1.4)
* Results from evaluations should be shared and published broadly to inform policy and program development and embed an evaluation culture. (Section 3.1.5)

**The most common types of evaluations used for assessment of health programs across Australia are:**

* Process evaluations using mixed methods data collection strategies. Pre-post designs were commonly reported by the systematic reviews. Ultimately, evaluation design should be considered on a case-by-case basis, using the better practice principles outlined in Section 3.1. (Section 3.2)

**Australian government agencies commonly:**

* Outsource their evaluations, whereas the private sector appears to both outsource evaluations and perform them in-house. There are otherwise no meaningful differences noted between government and private sector evaluations. (Section 3.3)

**The cost of conducting health program evaluations:**

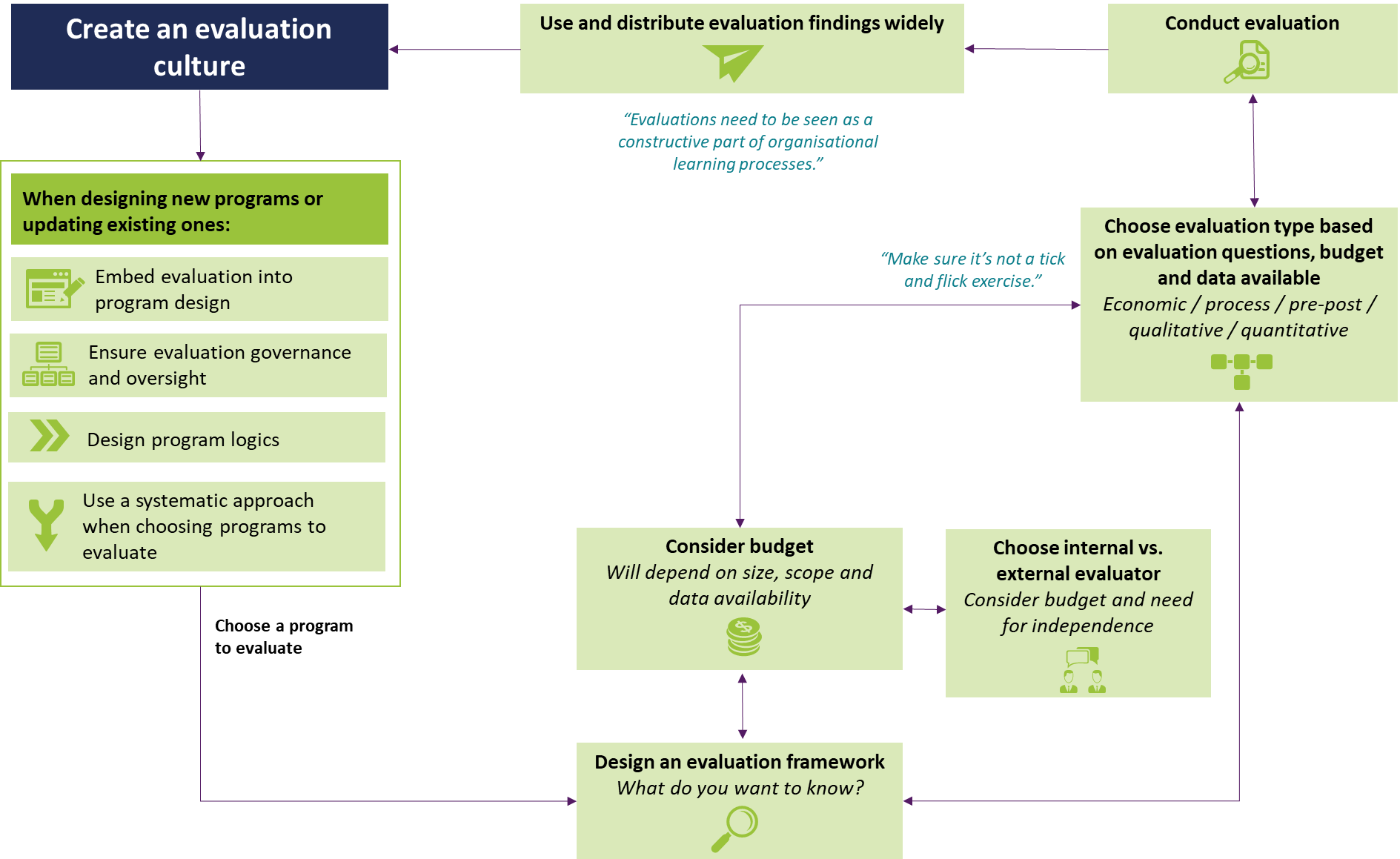
* Can vary greatly; while the NSW government has broadly suggested 10% of program costs, this number can diverge due to its scale, complexity in obtaining data, level of risk, intended use and significance to stakeholders and health outcomes. (Section 3.4)

**Changes in evaluation practice over time include:**

* Evaluations becoming increasingly becoming embedded into program design rather than being an external function. Continuous evaluation is becoming more common than single one-off evaluation. There is evidence from overseas that stakeholder engagement is increasing in evaluations, both in informing program design, and in collecting qualitative data on attitudes and behaviour. Evaluation remains an important but largely undervalued part of health programs. (Section 3.5)

In summary, there are no firm parameters to inform decisions on evaluation design, budget, and the choice between using external versus internal evaluators. Rather, better practice evaluations are based on a strong evaluation culture – whereby evaluative thinking and a continuous improvement approach is embedded into organisations, leading to high quality evaluations and thus high-quality program design and delivery (Figure 13). Key elements of this include designing program logics to inform future evaluations and distributing and sharing evaluation findings broadly.

***Figure 13: Better practice evaluation and evaluation culture***



1. References

1. **Australian Government Department of Veterans' Affairs.** Additional program information provided by DVA. 2022.

2. **Australian Government: The Treasury.** How to evaluate? [Online] 01 December 2021. https://evaluation.treasury.gov.au/toolkit/how-evaluate.

3. **Barton, Christian J et al.** *Program evaluation of GLA:D®Australia: Physiotherapist training outcomes and effectiveness of implementation for people with knee osteoarthritis.* s.l. : Elsevier Ltd, 2021.

4. **Barclay, Linda et al.** *Community integration programs and interventions for people with spinal cord injury: a scoping review.* 2020.

5. **Australian Government The Treasury.** What is evaluation? [Online] 27 April 2022. https://evaluation.treasury.gov.au/toolkit/what-evaluation.

6. **Blignault, Ilse et al.** *The value of partnerships: lessons from a multi-site evaluation of a national social and emotional wellbeing program for Indigenous youth.* s.l. : Australian and New Zealand Journal of Public Health, 2015.

7. **Browne, Jennifer et al.** *Food and nutrition programs for Aboriginal and Torres Strait Islander Australians: an overview of systematic reviews.* s.l. : Australian Health Review, 2018.

8. **Australian Government: The Treasury.** When to evaluate? [Online] 2021. https://evaluation.treasury.gov.au/toolkit/when-evaluate.

9. **Campbell, Brayden et al.** *A 'joint' approach: design, implementation and preliminary outcomes of an early conservative knee and hip osteoarthritis service in northern New South Wales, Australia.* 2020.

10. **Dabkowski, Elissa et al.** *An exploration into suicide prevention initiatives for mental health nurses: A systematic literature review.* s.l. : International Journal of Mental Health Nursing, 2021.

11. **Australian Government Department of Veterans' Affairs.** Our Wellbeing Landscape. 2018.

12. **Doran, Christopher M et al.** *Impact and Return on Investment of the Take Kare Safe Space Program - A Harm Reduction Strategy Implemented in Sydney, Australia.* 2021.

13. **Davies, Jenny et al.** *Developing an Australian health and aged care research agenda: a systematic review of evidence at the subacute interface.* 2016.

14. **KPMG.** *Evaluation and review of the National Suicide Prevention Leadership and Support Program.* 2021.

15. **Deans, C.** *Benefits and Employment and Care for Peer Support Staff in the Veteran Community: A Rapid Narrative Literature Review.* 2020.

16. **Australian Government Productivity Commission.** *A Guide to Evaluation under the Indigenous Evaluation Strategy.* 2020.

17. **Marsden, Elisabeth et al.** *A structure and process evaluation of the geriatric emergency department intervention model.* 2021.

18. **Geia, Lynore et al.** *Adolescent and young adult substance use in Australian Indigenous Communities: a systematic review of demand control program outcomes.* s.l. : Australian and New Zealand Journal of Public Health, 2018.

19. **Centers for Disease Control and Prevention.** CDC Coffee Break: Evaluative Thinking: Strategies for reflective thinking in your organisation. [Online] https://www.cdc.gov/dhdsp/docs/CB-June2018-508.pdf.

20. **Moreton, Sam et al.** *Economic and clinical outcomes of the nurse practitioner-led Sydney Adventist Hospital Community Palliative Care Service.* 2020.

21. **Lee, Xing Ju et al.** *Review of methods and study designs of evaluations related to clinical pathways.* 2019.

22. **Centers for Disease Control and Prevention.** Types of evaluation. *Program Performance and Evaluation Office (PPEO).* [Online] https://www.cdc.gov/std/program/pupestd/types%20of%20evaluation.pdf.

23. **Perceval, Meg et al.** *Evaluation of the SCARF Well-Being and Suicide Prevention Program for Rural Australian Communities.* 2019.

24. **Lim, Megan, S. C., et al.** *Reach, engagement, and effectiveness: a systematic review of evaluation methodologies used in health promotion via social networking sites.* s.l. : Health Promotion Journal of Australia, 2016.

25. **Edwards, Barry et al.** *Building research and evaluation capacity in population health: the NSW Health approach.* 2016.

26. **Rahja, Miia et al.** *Some gain for a small investment: An economic evaluation of an exercise program for people living in residential aged care.* 2020.

27. **Lokuge, Kamalini et al.** *Indigenous health program evaluation design and methods in Australia: a systematic review of the evidence.* s.l. : Australian and New Zealand Journal of Public Health, 2017.

28. **NSW Government Department of Premier and Cabinet.** NSW Government Program Evaluation Guidelines. [Online] January 2016. https://www.cveevaluation.nsw.gov.au/\_\_data/assets/pdf\_file/0007/771757/nsw-government-program-evaluation-guidelines.pdf.

29. **Scott, Theresa et al.** *Implementation and evaluation of a driving cessation intervention to improve community mobility and wellbeing outcomes for people living with dementia: study protocol of the 'CarFreeMe' for people with dementia program.* s.l. : BMC Geriatric, 2019.

30. **Newton, Danielle et al.** *A review of Evidence‐Based Evaluation of Measures for Assessing Social and Emotional Well‐Being in Indigenous Australians.* s.l. : Australian Psychologist, 2020.

31. **NSW Ministry of Health: Centre for Epidemiology and Evidence.** Commissioning Evaluation Services: A Guide. *NSW Health.* [Online] September 2019. https://www.health.nsw.gov.au/research/Pages/evaluation-guide.aspx.

32. **The George Institute of Global Health, EY, The University of Sydney .** *Evaluation of the Early Psychosis Youth Services Program.* 2020.

33. **Senanayake, Sameera et al.** *Impact and outcome evaluation of HealthPathways: a scoping review of published methodologies.* s.l. : The Royal New Zealand College of General Practitioners, 2021.

34. **NSW Ministry of Health, Centre for Epidemiology and Evidence.** Study Design for Evaluating Population Health and Health Service Interventions: A Guide. [Online] 2019. https://www.health.nsw.gov.au/research/Publications/study-design-guide.pdf.

35. **Tjalaminu, Mia et al.** *An Evaluation of the National Empowerement Project Cultural, Social, and Emotional Wellbeing Program.* s.l. : Journal of Indigenous Wellbeing, 2017.

36. **Snijder, Mieke et al.** *A systematic review of studies evaluating Australian indigenous community development projects: the extent of community participation, their methodological quality and their outcomes.* s.l. : BMC Public Health, 2015.

37. **Queensland Government Statistician's Office: Queensland Treasury.** Queensland Government Program Evaluation Guidelines, Second Edition. [Online] 2020.

38. **Trankle, Steven et al.** *The Nepean Blue Mountains Partners in Recovery Evaluation.* 2015.

39. **New Zealand Government: Social Policy and Research Unit (Superu).** *Making sense of evaluation: A handbook for everyone.* 2017.

40. **University of Queensland.** *Evaluation of the Pharmaceutical Benefits Scheme Subsidised Take Home Naloxone Pilot.* 2022.

41. **UK Evaluation Society.** *Guidelines for Good Practice in Evaluation .* 2018.

42. **University of Melbourne.** *National Suicide Prevention Trial.* 2020.

43. **Victorian Government Department of Health, Prevention and Population Health Branch.** Evaluation framework for health promotion and disease prevention programs. [Online] 2010. http://www.health.vic.gov.au/healthpromotion/evidence\_evaluation/cdp\_tools.htm.

44. **University of Queensland.** *Final report for the overarching evaluation of the National Support for Child and Youth Mental Health Program.* 2021.

45. **Volkov, Boris B. et al.** *A Checklist for Building Organisational Evaluation Capacity.* 2007.

46. **Australian Government The Treasury.** Evaluation in the Commonwealth (RMG 130). [Online] 23 May 2022. https://evaluation.treasury.gov.au/toolkit/commonwealth-evaluation-toolkit.

1. Desktop literature review summary

Table 16: Evaluation and Policy Documents

|  |  |  |
| --- | --- | --- |
| Author/Organisation (reference) | Document Title | Year |
| Australian Government Department of Finance (2) | How to evaluate? | 2021 |
| Australian Government Department of Finance (5) | What is evaluation? | 2022 |
| Australian Government Department of Finance (8) | When to evaluate? | 2021 |
| Australian Government Department of Finance (46) | Evaluation in the Commonwealth (RMG 130) | 2022 |
| Australian Government Department of Veterans' Affairs (11) | Our Wellbeing Landscape | 2018 |
| Australian Government Department of Veterans' Affairs (1) | Additional program information provided by DVA | 2022 |
| Australian Government Productivity Commission (16) | A Guide to Evaluation under the Indigenous Evaluation Strategy | 2020 |
| Centers for Disease Control and Prevention (19) | CDC Coffee Break: Evaluative Thinking: Strategies for reflective thinking in your organisation | 2018 |
| Centers for Disease Control and Prevention (22) | Types of evaluation. Program Performance and Evaluation Office (PPEO) | 2007 |
| Edwards, Barry et al. (25) | Building research and evaluation capacity in population health: the NSW Health approach | 2016 |
| New Zealand Government: Social Policy and Research Unit (Superu) (39) | Making sense of evaluation: A handbook for everyone | 2017 |
| NSW Government Department of Premier and Cabinet (28) | NSW Government Program Evaluation Guidelines | 2016 |
| NSW Ministry of Health, Centre for Epidemiology and Evidence (31) | Commissioning Evaluation Services: A Guide | 2019 |
| NSW Ministry of Health, Centre for Epidemiology and Evidence (34) | Study Design for Evaluating Population Health and Health Service Interventions: A Guide | 2019 |
| Queensland Government Statistician's Office, Queensland Treasury (37) | Queensland Government Program Evaluation Guidelines, Second Edition | 2020 |
| UK Evaluation Society (41) | Guidelines for Good Practice in Evaluation | 2018 |
| Victorian Government Department of Health, Prevention and Population Health Branch (43) | Evaluation framework for health promotion and disease prevention programs | 2010 |
| Volkov, Boris B. et al. (45) | A Checklist for Building Organisational Evaluation Capacity | 2007 |

Table 17: Program Evaluations

| **Author (reference)** | **Article Title** | **Year** | **In-house or commissioned** | **Program Description** | **Evaluation approach** | **Relevance to DVA and their programs** | **Relevant information** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Barton, Christian J et al. (3) | Program evaluation of GLA:D Australia: Physiotherapist training outcomes and effectiveness of implementation for people with knee osteoarthritis | 2021 | In-house | This evaluation assessed the implementation of an education program to support Australian physiotherapists to deliver guideline-based patient education and exercise-therapy to people with hip and knee osteoarthritis | Not stated | Community care and support  (relevant to Program 2.4) | * Evaluation focused on education approaches for healthcare practitioners, rather than a health program working directly with patients * Whilst not stated - hypothesised process & outcomes evaluation due to patient outcomes measures (outcomes) and established baseline data to examine effectiveness of implementation (process) * Survey response rates were low - authors would recommend, for similar programs, considering if mandatory post-training testing of health professionals is appropriate to improve response rates * A program logic was developed |
| Blignault, Ilse et al. (6) | The value of partnerships: lessons from a multi-site evaluation of a national social and emotional wellbeing program for Indigenous youth | 2015 | Commissioned | This evaluation assessed the first three years of a national program to improve the social and emotional wellbeing of Aboriginal and Torres Strait Islander young people aged 16-26 with a particular focus on depression, anxiety, violence and alcohol and other drug problems, by engaging young people and strengthening stakeholder and community responses to Aboriginal youth social and emotional wellbeing issues | Process & outcomes | Populations at increased risk (relevant to Mental and Social Health Programs) | * Whilst this is a youth-based program, it is relevant as it focuses on supporting children within a vulnerable population (which is a future area of focus for DVA, with Kookaburra Kids) * Evaluation combined elements of improvement-oriented open inquiry (drawing out lessons from the local communities, as well as through the program funder) and an audit review * Culturally appropriate questioning to gather qualitative research was essential and an enabler for success for the evaluation * Authors noted that capturing the impact of community strengthening initiatives is not easy as a long-term view is required, especially in disadvantaged communities where circumstances can change rapidly * Program logic developed with input from community and stakeholders |
| Campbell, Brayden et al. (9) | A 'joint' approach: design, implementation and preliminary outcomes of an early conservative knee and hip osteoarthritis service in northern New South Wales, Australia | 2020 | Commissioned | Community health-based conservative OA joint management service in the Tweed Valley | Process & outcomes | Community care and support  (relevant to Program 2.4) | * Program logic developed by the funders of the program * Process evaluation was qualitative and the outcomes evaluation was quantitative * For further information see Deloitte Access Economics evaluation: <http://eih.health.nsw.gov.au/__data/assets/pdf_file/0008/570869/OACCP-evaluation-feb-2015.pdf> |
| Doran, Christopher M. et al. (12) | Impact and Return on Investment of the Take Kare Safe Space Program - A Harm Reduction Strategy Implemented in Sydney, Australia | 2021 | Commissioned | Evaluation of the Take Kare Space Program - a harm reduction program implemented to address alcohol-related violence and disorder in three locations in Sydney's night-time economy between 2014 - 2019 | Process, outcomes & economic | Populations with mild - moderate needs (relevant to Mental and Social Health Programs) | * Process, outcomes and economic indicators were examined * A mixed-methods approach was embedded into the evaluation framework, combining both qualitative and quantitative methods |
| KPMG (14) | Evaluation and review of the National Suicide Prevention Leadership and Support Program | 2021 | Commissioned | This program funds a range of activities designed to reduce deaths by suicide nationally | Process | Populations at increased risk (relevant to Mental and Social Health Programs) | * This evaluation included mixed methods approach, data analysis of financial and performance data, income statement budgets, work plans, audited financial statements; semi-structured interviews, and documentation review (other documents provided by partners) |
| Marsden, Elisabeth et al. (17) | A structure and process evaluation of the Geriatric Emergency Department Intervention model | 2021 | Commissioned | The Geriatric Emergency Department Intervention (GEDI) is a nurse-led, physician-championed innovative model that aims to streamline service delivery by maximising the provision of care for frail older adults during their ED journey. GEDI nurses were based in ED, and worked as a supplementary sub-specialty team in ED facilitating the primary ED nurse and physician care. They helped identify patients on arrival in ED and they performed rapid patient-centred, focussed comprehensive geriatric assessments and fast-tracked diagnostic processes and referrals | Process | Hospital services (relevant to Program 2.2) | * Evaluated service structures and organisational structures; aim was to describe and evaluate the key structures (e.g. staff, funding, resources) and processes (e.g. communication skills, holistic care) required to achieve the favourable outcomes of the model * Recognises the need to identify the structures and processes necessary for transferability of this model into other health services * Descriptive qualitative study utilising Donabedian’s framework for evaluating health services, 24 semi-structured interviews were done * An a priori framework guided interview questions and analysis * Two analytical frameworks used: Firstly, Donabedian’s differentiation of structures and processes and secondly, Irvine et al.’s Nursing Role Effectiveness Model (NREM) which conceptualises nurses’ contributions to healthcare |
| Moreton, Sam et al. (20) | Economic and clinical outcomes of the nurse practitioner-led Sydney Adventist Hospital Community Palliative Care Service | 2020 | Commissioned | This program is a nurse practitioner led Sydney Adventist Hospital Community Palliative Care Service (SanCPCS) | Outcomes evaluation | Community care and support (relevant to Program 2.4) | * This article does include the economic impact of this program, including the change in patient costs before and after service expansion, as well and changes in hospital admissions * To calculate estimated cost, the number of hospital admissions, LOS and type of admission were assessed for each patient. The Australian Refined Diagnosis Related Groups (AR-DRGs) system was used to estimate day-specific acute care hospital costs |
| Perceval, Meg et al. (23) | Evaluation of the SCARF Well-Being and Suicide Prevention Program for Rural Australian Communities | 2019 | Commissioned | This program is a well-being and suicide prevention education workshop, SCARF (Suspect, Connect, Ask, Refer, Follow-Up) developed for Australian farming and rural communities | Not stated | Community care and support (relevant to Program 2.4) | * The Literacy of Suicide Scale (LOSS), Stigma of Suicide Scale (SOS), and items assessing confidence to assist others were administered immediately before and after the workshop, and at 3-month follow-up [to capture both short- and longer-term changes in suicide literacy and stigma]. The Warwick Edinburgh Mental Wellbeing Scale was given immediately before and at 3-month follow-up * Data were analysed using linear mixed-effects regression |
| Rahja, Miia et al. (26) | Some gain for a small investment: An economic evaluation of an exercise program for people living in residential aged care | 2020 | In-house | This 12 week exercise physiology (EP) program was delivered by an accredited exercise physiologist (AEP) and designed for frail older people with cognitive impairment living in RACF in South Australia | Cost utility analysis (economic analysis) | Community care and support (relevant to Program 2.4) | * Evaluated the outcome of quality-adjusted life years (QALY) gained * Conducted a preliminary analysis of cost and quality of life (QoL) outcomes of the program. The outcomes were then converted to utilities and expressed as cost per QALY gained * Analysis was completed from a health service provider perspective, meaning only costs directly related to service providers delivering the program were included * Given short timeline of program, expected costs/benefits to occur simultaneously (no discounting applied) |
| Scott, Theresa et al. (29) | Implementation and evaluation of a driving cessation intervention to improve community mobility and wellbeing outcomes for people living with dementia: study protocol of the ‘CarFreeMe’ for people with dementia program | 2019 | In-house | CarFreeMe is an intensive program delivered by a trained health professional that addresses practical and emotional needs relevant to driving cessation. It is a support- and education-based intervention targeted at people living with dementia and their care partners who are managing driving cessation. | Process | Community care and support (relevant to Program 2.4) | * A cluster randomized controlled trial was done to evaluate the effectiveness of the program. Evaluation took place pre-intervention, immediately following, and three months post-intervention. Clusters were randomized to either intervention or usual treatment * Data was analysed using random effects models with maximum likelihood estimation. All analyses were conducted at the individual and aggregated cluster levels using generalized estimating equations to adjust for clustering effects within geographical location, and adjusting for intra-cluster correlations between individuals, and according to intention to treat principles. * Qualitative analyses will include analysis of the semi-structured interviews with participants. Coding will be conducted and verified by minimum three members of the research team. The group will convene to review themes until consensus is reached. * A process evaluation will explore the implementation, receipt, and setting of the intervention and help in the interpretation of the outcome results. In trials, where the “same” intervention may be implemented and received in different ways, process evaluations can provide greater explanatory power and understanding of the generalisability of the intervention It will contribute valuable information to understanding optimal processes, for example timing and number of sessions, in determining a feasible translation strategy |
| The George Institute of Global Health, EY, University of Sydney (32) | Evaluation of the Early Psychosis Youth Services Program | 2020 | Commissioned | The Early Psychosis Youth Services Program provided integrated early intervention treatment for young people aged 12-25yrs at risk or experiencing psychosis | Process, Outcome, Economic | Populations at increased risk (relevant to Mental and Social Health Programs) | * Mixed method design was used which enabled a multi-pronged, methodological approach to help answer objectives. The sources of data included: consultation with local stakeholders with direct experience of the EPYS Program; client and family interviews; a family and carer survey; program specific data including the headspace Application Platform Interface (hAPI) Minimum Data Set (MDS) and workforce and financial data; case studies of usual care; comparative data including: (1) a like-service control comparison using the comparative service cohort; and (2) an ecological counterfactual using NSW Health and WA Health data;1 and relevant literature and program documents * A program logic was included |
| Tjalaminu, Mia et al. (35) | An Evaluation of the National Empowerment Project Cultural, Social, and Emotional Wellbeing Program (CSEWP) | 2017 | Commissioned | This evaluation assessed from participant interviews and Stories of Most Significant Change (SMSC) if, and how, the CSEWB program in Indigenous communities of Queensland contributed to strengthening the cultural, social, and emotional wellbeing of participants, their families and communities. The 6-week program focused on strengthening the different domains of cultural, social, and emotional wellbeing (of individuals, families, and the community) and involves a number of activities that have a community-wide focus. | Stories of most significant change | Community care and support (relevant to Program 2.4) | * This qualitative post program evaluation identified the perceived benefits of the program for individual participants, their families, and communities. * Stories of most significant change (SMSC) and interviews (with open ended questions) were collected. The interviews informed case studies * SMSC are valuable in measuring outcomes of social change programs; suitable for researchers who have regular contact with participants, who wish to collect data on impact rather than outputs |
| Trankle, Steven et al. (38) | The Nepean Blue Mountains Partners in  Recovery Evaluation | 2015 | Commissioned | This evaluation assessed the implementation of the 'Partners in Recovery' program whose goal is to better support people with mental illness and their carers across the country. A problem logic model has been developed, and progress measured using mixed method research | Not stated | Community care and support (relevant to Program 2.4) | * This program looked at improving care through coordination and integration of services and support across sectors * Impacts envisaged as a result of this program included enhanced local community health and wellbeing, and  integrated and co-ordinated health services * Program Logic Model described intended inputs, activities, outputs, outcomes and impacts of program; indicators were assigned to each, and progress measured towards each indicator measured with mixed methods approach * Mixed methods included collection and analysis of Australian and international literature; program policy, operational and reporting docs; quantitative and qualitative survey data |
| University of Queensland (40) | Evaluation of the Pharmaceutical Benefits Scheme Subsidised Take Home Naloxone Pilot | 2022 | Commissioned | The Australian Government funded a Take Home Naloxone (THN) Pilot as part of the Pharmaceutical Benefits Scheme (PBS), allowing people at risk of experiencing or likely to witness an opioid overdose to access naloxone without a prescription, at no cost to themselves, and from a range of pharmacies and other approved sites in New South Wales, South Australia and Western Australia. | Outcome | Medicines use or PBS utilisation (relevant to program 2.3) | * The Evaluation used a mix of primary data collection activities to address the key evaluation questions * Both quantitative and qualitative data were collected from people who participated in the Pilot as consumers (recipients of THN) and front-line staff of sites where naloxone was provided (access site staff) * Structured consultations with sector representatives and systematic analyses of Pilot documentation were also undertaken * Analyses of administrative and research data relating to the supply and use of opioids and naloxone were also conducted |
| University of Melbourne (42) | National Suicide Prevention Trial | 2020 | Commissioned | The National Suicide Prevention Trial was announced by the Australian Government in 2016 and aimed to gather evidence and further understanding of what strategies are most effective in preventing suicide at a local level and in at-risk populations. Twelve trial sites commissioned a range of evidence based and innovative interventions guided by multi-component models. | Process and outcome | Populations at increased risk (relevant to Mental and Social Health Programs) | * The evaluation used a mixed-methods design drawing on quantitative and qualitative data from a range of sources   + review of key documents   + stakeholder consultations   + observational / participatory data   + aftercare service user data   + community based activities data * Evaluation framework provided by Department of Health, program logic included    + national epidemiological and service use datasets   + systems modelling and simulation |
| University of Queensland (44) | Final report for the overarching evaluation of the National Support for Child and Youth Mental Health Program | 2021 | Commissioned | The National Support for Child and Youth Mental Health Program (the Program) aims to improve mental health outcomes for children and young people, commencing with the early years and going through to adolescence, by providing targeted grants for workforce and education activities that will build capabilities aligned to the Program objectives | Process, Outcome, Economic | Populations at increased risk (relevant to Mental and Social Health Programs) | * This Overarching Evaluation used a suite of complementary qualitative and quantitative research activities to address the Evaluation Questions. The main research methods used included: * Design and administration of the National Support Network Survey: a national survey of Educators and Practitioners as the key target audiences for Be You and Emerging Minds, respectively   + An Integrated Data Analysis: a secondary analysis of multiple existing datasets appropriate to child and youth mental health   + A Value for Money assessment, including estimates of funding inputs from the Commonwealth Government and alternative scenarios generated from qualitative survey and interview data   + Community Case Studies: interviews and focus groups conducted with key stakeholder groups in four communities in Queensland and Western Australia, including Educators, Practitioners (clinical and non-clinical), and parents/carers of young people   + Document analysis of evaluation reports from the two Program initiatives: Be You and Emerging Minds   + Key Program Informant Interviews with Be You and Emerging Minds staff |

Table 18: Systematic Reviews

| **Author** | **Article Title** | **Year** | **Key Findings** | **Relevance to DVA and their programs** |
| --- | --- | --- | --- | --- |
| Barclay, Linda et al.  (4) | Community integration programs and interventions for people with spinal cord injury: a scoping review | 2020 | * Of the evaluations reviewed, there was great variation in the study designs and methods used, including pilot studies (n=2), mixed methods evaluation (n=2), single-site randomised controlled trials (n=3), and randomised single-arm study design (n=1) * Evaluating multi-disciplinary programs that deliver a range of interventions is challenging, as such programs need to be personalised to meet the needs of individual patients * Some studies reviewed highlighted that qualitative feedback was able to gauge patient and family satisfaction, however, no significant change was identified in quantitative outcomes in one of the randomised control trials * Acknowledgement by the authors that scoping reviews do not typically assess the quality of included articles, therefore results are more challenging to interpret and limits uptake into policy and practice | Community care and support (relevant to Program 2.4) |
| Browne, Jennifer, et al.  (7) | Food and nutrition programs for Aboriginal and Torres Strait Islander Australians: an overview of systematic reviews | 2018 | * Rigorous mixed-method approaches can enhance the evaluation of health promotion programs * The social, cultural and geographical diversity of Aboriginal and Torres Strait Islander communities in Australia meant that a one-size-fits-all approach is unlikely to be successful * In order to improve health outcomes, other social and cultural determinants of health must be addressed alongside nutrition | Populations at increased risk (relevant to Mental and Social Health Programs) |
| Dabkowski, Elissa et al.  (10) | An exploration into suicide prevention initiatives for mental health nurses: A systematic literature review | 2021 | * Articles included in the review consisted of quantitative (n=10), mixed methods (n=2), and qualitative studies (n=2) * Frequently used study design within the selected papers was a pre- and post-test evaluation methodology, with data collected immediately before and after the intervention * Limitations of the final data set included self-reporting and an absence of long-term follow-up * Some authors within the articles noted that participants were motivated to attend the intervention, which potentially skews results * Some articles discussed the lack of control groups to their intervention may have been a limitation to findings * Thematic analysis using focus groups was the predominant choice for qualitative research | Prevention & Early intervention (relevant to Mental and Social Health Programs) |
| Davis, Jenny et al.  (13) | Developing an Australian health and aged care research agenda: a systematic review of evidence at the subacute interface | 2016 | * The different studies used varying research methods, including: prospective comparisons with matched controls, prospective observational cohort study with baseline and repeated-measures follow-up, two group pathways, observational case series, prospective non-random with matched controls, pilot program, Quasi-RCT, Observations of records and interviews, Blinded RCT, RCT, written surveys and semi-structured interviews | Community care and support (relevant to Program 2.4) |
| Deans, C (15) | Benefits and Employment and Care for Peer Support Staff in the Veteran Community: A Rapid Narrative Literature Review | 2020 | * Some of the single program interventions reviewed also had preliminary evaluation, including some 12-month follow up * A synthesis of information from the general mental health peer field (where factors not relevant to a veteran environment were removed) indicated external independent evaluation was part of the 'best practice framework' * The process for evaluation of interventions remains in its infancy, with several qualitative studies, few control group measurements and many evaluations of low power or quality | Community care and support (relevant to Program 2.4) |
| Geia, Lynore et al.  (18) | Adolescent and young adult substance use in Australian Indigenous communities: a systematic review of demand control program outcomes | 2018 | * Acknowledgement from the author that some programs on adolescent and young adult AOD use in Indigenous communities may not be ready for evaluation or may be inadequately funded. The need to understand the broader drives of effective program delivery (e.g. need for self-determination and to consider local cultural context) is important. Implementation issues need to be addressed before outcome evaluations become possible. * One study noted that it is only possible to establish program outcomes when valid and reliable tools are used to assess need, and that baseline data are required to document the extent to which change over time occurs. * Of the studies reviewed, all used outcomes measures. Varying study techniques were used. | Populations at increased risk (relevant to Mental and Social Health Programs) |
| Lee, Xing Ju et al.  (21) | Review of methods and study designs of evaluations related to clinical pathways | 2019 | * The predominant study design used in the evaluations identified was the pre-post study design which identifies and collects data for a control group before the intervention implementation and compares the data to the intervention group data (72.7%). The second most commonly used study design was the concurrent case-control design where data for both the control and intervention groups were collected in the same time period (14.5%). * Almost all data collected individual patient data directly (96.4%) * Most of the statistical methods used in the investigations can be broadly classified into one of the four categories: comparison of group means, omnibus tests, comparison of group distributions and estimation of covariate effects. * The majority of economic evaluations identified were cost-consequence analyses where the costs and outcomes of the interventions were reported separately. | Medical services (relevant to Program 2.1) |
| Lim, Megan S. C. et al.  (24) | Reach, engagement, and effectiveness: a systematic review of evaluation methodologies used in health promotion via social networking sites | 2016 | * Two main evaluation approaches were used: 1) 'closed' (n=23) used a traditional research design (e.g. RCT, cross-sectional, pre-post), formal recruitment procedures (e.g. F2F w/ consent procedures) and had a captive or target group of participants 2) 'open' (n=19) evaluated the intervention in a real-world setting and allowed any social network users to follow and interact with the content without any formal enrolment procedures and often users were unaware they were being evaluated * Studies using a closed designed were categorised by more rigorous evaluation design, often emphasising assessment of effectiveness * Evaluations of effectiveness included RCTs (which are considered gold-standard in study design), however, majority of RCTs relied on subjective measures, particularly self-reported surveys * Low response and retention rates were key challenges outlined - however, are noted to be an ongoing challenge in eHealth   Recommendations:   * Evaluations can balance rigour and real-world application * Where RCTs are not possible, consider quasi-experimental designs * Consider multiple aspects of the intervention in evaluations – reach, engagement, and effectiveness are all important * Avoid using cross-sectional surveys without a comparison group to measure effectiveness * When assessing reach, think beyond raw numbers of likes and followers, consider denominators and demographics or characteristics of users * Use mixed-methods evaluation design where possible, especially for measuring engagement * Consider a variety of measures to determine effectiveness – ecological, objective, qualitative and quantitative * Objective or biological outcomes are important to enable assessment of effectivenes and impact and should be considered in the planning phase of the intervention. | Prevention & Early intervention (related to Mental and Social Health Programs) |
| Lokuge, Kamalini et al.  (27) | Indigenous health program evaluation design and methods in Australia: a systematic review of the evidence | 2017 | * Of the evaluations reviewed, 82.2% included a quantitative component, with 49.2% using only quantitative data, and 33.1% using both quantitative and qualitative data * The most common evaluation design was a before/after comparison (30.5%). 7.6% used experimental design; this included RCTs and cluster-RCTs * More than half the evaluations (56.8%) measured a health service delivery or process outcome; 33.1% measured a health or health risk factor outcome, and 10.2% measured both * All studies that employed an experimental design measured a health outcome or risk factor, rather than a health service delivery or process outcome only * Of the non-experimental evaluations, 61.5% measured a health service delivery or process outcome, 28.4% measured a health or health risk factor outcome, and 10.1% measured both * Quantitative data provide a measure of health impact on health programs; qualitative data enables participative and collaborative evaluation and are valuable to contextualise and provide culturally relevant inferences about a program's utility and impact - therefore, a mixed methods approach utilising both quantitative and qualitative data is likely to be the best approach to conducting program evaluations * Experimental designs, such as RCTs and cluster-RCTs, are only possible when considered during program development and implementation, and if relevant to and supported by participating communities (they may also not be appropriate for every setting, particularly in the case of complex public health interventions or when the evaluation is conducted with a small sample or in one setting) | Populations at increased risk |
| Newton, Danielle et al.  (30) | A review of Evidence‐Based Evaluation of Measures for Assessing Social and Emotional Well‐Being in Indigenous Australians | 2020 | * This article systematically reviews available measures of social and emotional well-being (SEWB) after an intervention, classifying them in terms of the evidence base that exists to support their use * It is concluded that there is an ongoing need to develop psychometrically sound, comprehensive, culturally appropriate measures to operationalise Indigenous SEWB at a population health, programme evaluation, and clinical level * Review evidence of instruments/measures to assess social and emotional wellbeing (SEWB) and how it changes over time as a result of intervention * This study adopted the APA Division 54 (Society of Paediatric Psychology) assessment task force steering committee (SPP-ATF) to classify assessment instruments (measures) as “evidence-based”, categorising them into three possible evidence-based categories: (a)well-established assessment, (b) approaching well-established assessment; and (c) promising assessment * Measures investigated include measures of psychological distress/psychopathology, quality of life and empowerment * Well-established assessment: The measure must have been presented in at least two peer-reviewed articles by different investigators or investigatory teams. Sufficient detail about the measure to allow critical evaluation and replication. Detailed statistical information indicating good validity and reliability in at least one peer-reviewed article. | Populations at increased risk (relevant to Mental and Social Health Programs) |
| Senanayake, Sameera et al.  (33) | Impact and outcome evaluation of HealthPathways: a scoping review of published methodologies | 2021 | * Twenty-one studies were included in the final review * The HealthPathways programme is an online health information system used mainly in primary health care to promote a consistent and integrated approach to patient care. * The aim of this study is to perform a scoping review of the methodologies used in published impact and outcomes evaluations of HealthPathways programmes * The review included qualitative, quantitative or mixed-methods evaluations of the impact or outcome of HealthPathways * ‘Increased awareness and use of HealthPathways’ was the most frequent programme aim evaluated. Quantitative and qualitative research methodologies, as well as prospective and retrospective data collections, have been adopted to evaluate the impact and outcome of HealthPathways * Qualitative research methods are also best placed to gather the in-depth insights required to understand the barriers and facilitators to program use, as this information is currently not well understood by other means. * Data collection and analysis of qualitative studies requires expertise and can be time consuming, and findings may not be transferable to other contexts. Furthermore, the perspectives of all relevant stakeholders may not be adequately captured during qualitative inquiries * This review was limited to impact and outcome evaluations * Year of publication, study population, study setting, study design, data collection (prospective, retrospective, or both), data source (hospital records, interviews, google analytics, focus groups, surveys), programme aims evaluated, indicators, evaluation design (descriptive, qualitative, mixed method, pre-post study), analytical method (descriptive statistics, thematic analysis, logistic regression), and results were extracted from each publication or report. * The impact and outcomes framework for this scoping review was developed using the six aims identified in a South Australian HealthPathways report and one additional aim was performed in accordance with the Preferred Reporting Items for Systematic Reviews identified during the process of this review: components related to awareness, usage, improvement in patient care and patient experience, and providing best value for money * It describes the strengths, limitations, and gaps of the current HealthPathways evaluation literature * Review was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension for scoping reviews | Medical services (relevant to Program 2.1) |
| Snijder, Mieke et al.  (36) | A systematic review of studies evaluating Australian indigenous community development projects: the extent of community participation, their methodological quality and their outcomes | 2015 | * Thirty one evaluation studies of community development projects were identified * Partnerships between researchers, community members and service providers have great potential to improve methodological quality and community participation when research skills and community knowledge are integrated to design, implement and evaluate community development projects * The methodological quality of studies evaluating Australian Indigenous community development projects is currently too weak to confidently determine the cost-effectiveness of community development projects in improving the health and wellbeing of Indigenous Australians * Qualitative and quantitative study components were assessed against appropriate criteria the Dictionary for Effective Public Health Practice Project Quality Assessment tool was used to assess the methodological quality of quantitative components and an adaptation of the qualitative study appraisal tool, developed by Long and Godfrey was used for the qualitative study components * Levels of community participation fluctuate across community development project phases: moderate in the Diagnosis and Development phases, high in the Implementation phase, but low in the Evaluation phase methodological quality of studies evaluating * Methodological quality of studies evaluating Australian Indigenous community development projects is too weak to confidently determine the cost-effectiveness of these projects in improving the health and wellbeing of Indigenous Australians * Partnerships combining researchers’ expertise and community members’ skills and knowledge have great potential to improve methodological quality and community participation in Indigenous community | Community care and support (relevant to Program 2.4) |

1. Key informant interviews summaries

Table 19. Key informant interview 1

|  |  |  |
| --- | --- | --- |
| **EY Executive Director, Technology Consulting, Data & Analytics (United States)** | | |
| **Interview date and time** | | Tuesday 24 May 2022, 8:00am – 9:00am AEST |
| **Theme** | **Key Findings** | |
| **Veterans’ health program evaluations landscape in the United States** | The US Department of Veterans Affairs (VA) and other government agencies in the United States (US) typically conduct three types of evaluations:   1. Evaluations with a performance management bent, used to assess whether specified outcomes have been achieved for the population at large. These tend to be managerial in nature. 2. Specialised evaluations for innovative programs, involving novel therapeutics or procedures, and therefore focused on a small cohort, usually in partnership with research institutions (where the novel therapeutic or procedure has been developed) 3. Evaluations carried out by the United States Congress Government Accountability Office, who conduct evaluations on various government departments and executives to understand the progress and performance of a program, this is to determine if programs continue | |
| **Evaluation methodologies used** | For the type of evaluations described above:   1. Program outcomes are defined, and normative values set. Time-trended data are collected and compared against both the normative value and identified external benchmarks (of good practice). 2. Methodologies are bespoke and complex, due to the need to determine a causative effect of the novel treatment and generally involve experimental methods comparing the treatment group to a comparison cohort e.g. usual care. Comparison cohorts must be comparable to treatment cohorts e.g. demographics, co-morbidity 3. These evaluations generally involve interviews and reviews of administrative data, to produce recommendations for the program | |
| **Key success factors** | * When looking at the first type of evaluation, establishing the goals and outcomes of a health program, the following questions must be considered: Is what you are intending to measure *meaningful*? Can the goal be measured in a *scientific* manner? Is the data *reliable*? Is it *feasible* to measure and evaluate? * Getting political agreement on the outcomes (and also on how they should be operationalised) for programs is also critical – this can be challenging, especially when numerous stakeholders are involved but sets the foundations for buy in when results arrive. Buy in especially at the higher levels of the organisation is crucial * Evaluation outcomes need to be tailored and specific for the population in mind, in this case veterans e.g. with the increasing number of women that are now serving on the front line, clinical aspects such as pregnancy and giving birth need to be considered, this wasn’t the case previously. Another example – specific injuries in context of particular conflicts e.g. TBI | |
| **Common pitfalls and challenges** | * A common challenge is the definition of wellness, and how this is articulated as a specific and measurable indicator for the population in question. * Agreement on what the goals and outcomes of the program are, and how they should be operationalised can be the most time-consuming activity in an evaluation | |
| **Government agency evaluations compared to private sector evaluations** | * Government agencies in the US often conduct evaluations in-house, particularly for the first type of evaluation. However, evaluations for novel programs may be fully commissioned or within a hybrid model, to access more specialised evaluation capabilities, given the more complicated methodologies * In the private sector, defining desired outcomes may be less complex and time consuming in comparison to the government agencies, e.g. there may a clear commercial aspect driving evaluation. This can be a reflection of the audience for the evaluation e.g. shareholders vs the US government on behalf of the public (where goals and outcomes may be more broadly defined) | |
| **Change in methods over time** | * In the US, ‘rapid cycle evaluations’ were introduced 5 – 10 years ago, as a means to provide quicker answers as to whether programs were performing or heading in the intended direction, and can be undertaken while programs are being implemented. These types of evaluation are less definitive with the advantage of speed is offset by the confidence levels in the results, which is ultimately driven by the types of questions being asked the level of confidence required by the evaluation’s audience * In the US, evaluations have only been conducted in the last 30 years, previously, they did not occur in the systematic way that they do now – with the effect that evaluations place greater accountability on program administrators to clarify what they want to achieve and a potential flow-on impact to program design * Stakeholder engagement has become increasingly important, especially in achieving political agreement on program goals | |
| **Costs of evaluations** | * N/A | |

Table 20. Key informant interview 2

|  |  |  |
| --- | --- | --- |
| **EY Partner, Strategy & Transactions (United States)** | | |
| **Interview date and time** | | Tuesday 24 May 2022, 8:00am – 9:00am AEST |
| **Theme** | **Key Findings** | |
| **Evaluations conducted within the US Department of Veterans’ Affairs** | * The US Department of Veterans’ Affairs (VA) is set up across three departments:  1. Health Administration (responsible for delivering health services and is the largest health provider in the US) 2. Benefits Administration (responsible for compensation, disability, pensions, education and other benefits that veterans may be entitled to) 3. Memorial Affairs (responsible for managing internments and running cemeteries)  * In 2018, an Evidence-Based Policy mandate was put in place to enforce every federal agency (including the VA) to look at outcomes-based measures and ‘create a learning agenda’. This involved organisations putting in place a plan to improve and progress thinking about their path to measuring programs and quantifying the benefits of their programs. The move to more meaningful measures and outcome-driven evaluations allows the Agency to understand and make better investment decisions. * The Health Administration and Benefits Administration have their own internal functions to perform health program evaluations. | |
| **Key challenges for to the US Department of Veterans’ Affairs in undertaking better practice evaluations** | * There currently is no Performance Evaluation Office within Health Administration, and therefore there is an opportunity to establish proper governance and oversight to greater understand evaluation data to inform decision making. * There is an opportunity for decisions to be made based on more robust data and in coordination with broader leadership within the Agency. * In the US, the government is fast-moving, and legislation is frequently changed. Evaluations are often unable to keep up with this pace of change, and cannot completely focus on program quality, or improvement opportunities. * There is an Inspector General for every federal Agency in the US who is an internal auditor and acts in an administrative capacity and conducts reviews which are often more critical in nature than evaluations. * Each federal Agency also has a government accountability office (GAO) which provides independent assessments of how Agencies operate. This provides an opportunity for ‘high risk’ programs and aspects of the Agency to be identified, contributing to greater transparency and scrutiny to drive improvement. | |
| **Key to successful evaluations** | * Key enablers of successful evaluations of US veterans’ health programs are governance structures and legislation. They support evaluation activities and help embed evaluative thinking throughout the organisation. | |

1. Case study summaries

Table 21: Case Study I

|  |  |  |
| --- | --- | --- |
| **Program evaluation unit, NT Department of Treasury and Finance** | | |
| **Interview date and time** | | Monday 30 May 2022, 4:00pm – 4:30pm AEST |
| **Theme** | **Key Findings** | |
| **Types of evaluations being conducted in NT** | * NT is early in the evaluation journey with work arising out of a fiscal policy report in 2019 which identified the need for budget repair over time, especially in the medium term. Three recommendations related to evaluation recognising the need to know the evidence base for a program, to inform decisions around resource allocations. * Previously evaluations in NT were more ad-hoc, or in response to crises. * The office has now established a systematic approach, and hope to drive consistent standards across the agencies. * Have developed a program evaluation framework which sets out expectations on evaluation type and includes guidance on how to tier evaluations (developed using guidance from other governments, in particular NSW) – recognise that from a budget perspective, they don’t need to evaluate everything; rather they need to identify appropriate programs and prioritise accordingly. * The program evaluation framework outlines 3 types of evaluations.  1. Process – most of the evaluations being conducted now are these given how early they are in their evaluation journey. 2. Outcome – to understand medium-term outcomes. 3. Impact – to understand long-term outcomes and causation or attribution/understanding what would happen in the absence of the program.  * Currently do not have the capability to undertake randomised controlled trials. | |
| **Key learnings for setting up an evaluation team** | * One of their key learnings was to implement a ‘program master list’.   + This came out of the 2016 audit of the NSW Program Evaluation Unit which critiqued how they organisationally arranged evaluations.   + This register is very important, particularly due to large staff turnover – having the information in a central area supports overarching governance. * Another key learning was to ‘close the loop’ on evaluations.   + NT need to inform Cabinet of how they intend to prioritise evaluations, it is also important to report back to Cabinet the following year to update them on what has been evaluated, the main findings and what the Agency’s response was. This helps encourage evaluations to be conducted. | |
| **Major challenges faced** | * The main challenge identified was around developing a fit-for-purpose program master list. NT originally borrowed a template from a NSW Agency and found that people struggled to complete the form. The template was since updated to respond to feedback and a more streamlined and easy process was formed. | |
| **In-house vs commissioned** | * NT both commission evaluations and uses in-house expertise to conduct evaluations. * Most of the program evaluations conducted have been commissioned as they are trying to build their evaluation capability. * NT have found the Australian Evaluation Society workshops to be particularly helpful for building evaluation capability. | |
| **Key success factors to make the process streamlined, efficient and cost-effective** | * Changing the Cabinet Submission systems and templates to provide more information required to request funding from Cabinet has been quite successful. * Having a program evaluation toolkit that provides guidance on needs assessments and relevant evidence banks and data warehouses has made it easier for people as they are aware of the evidence available and how to access it. | |
| **How best to leverage existing materials on evaluations** | * NT acknowledged the usefulness of the Australian Public Sector Evaluation network, which has a SharePoint site where members can upload templates and good examples of evaluations. It is noted to be slightly inactive at times as it is run by volunteers. * The importance of coordinating evaluations was also highlighted – government organisations can work together to share evaluation schedules and see how to coordinate their evaluations accordingly to avoid duplications and gaps and to share lessons learned. | |
| **Approach to supporting Agencies to conduct evaluations** | * NT have provided guidance in a toolkit to encourage Agencies to think about how to approach evaluations, however, have left the design of evaluations up to Agencies, who are more familiar with the context. | |

Table 22: Case Study II

|  |  |  |
| --- | --- | --- |
| **Strategic advisor to the NSW Ministry of Health** | | |
| **Interview date and time** | | Monday 06 June, 3:15pm – 3:45pm AEST |
| **Theme** | **Key Findings** | |
| **Types of evaluations and approaches** | * Typically two primary approaches to evaluations; program logic and helicopter. Program logic approach is deal, better at understanding the components at play that may need to be amended or which are working well. Helicopter evaluation more traditional and less useful in identifying areas needing improvement. * Program logic   + Most useful for programs with outcome timelines (e.g. short and long term outcomes).   + Can help pick out faults at play; e.g. not achieving desired outcomes due to lack of optimisation, basic model or design problems.   + Help clarify assumptions of the program.   + Help guide in identifying and achieving desired outcomes over time. * Theories of change is another useful methodology; however it is less useful when there are time dependent outcomes. | |
| **Purpose of evaluation** | * Evaluations can serve as a method to better understand where faults lie, identifying potential causes of unfulfilled outcomes. * Important to recognise any underlying assumptions in order to identify potential fault areas. | |
| **Embedded evaluation into program implementation** | * Applied evaluation approach, which involves embedding evaluation with program rollout or expansion can aid in ensuring evaluation results have practical use. * There has been a major shift towards this method in the past 10 years, rather than a separate external evaluation process. Now evaluation is commonly a significant part of continuous improvement cycles of program implementation, rather than a ‘petri dish’ one off evaluation approach.   E.g. giving feedback from evaluation to participants to help improve program implementation. | |
| **Importance of comparison** | * Even if exact RCT framework is not possible, important to still have a baseline comparison of some sort.   E.g. if a health program shows great results but the baseline comparison group does too, questionable benefits.   * Important to separate program effects from natural progression, clarify program benefits. | |
| **Collecting useful data** | * Mixed methods is an ideal approach; the a ‘single number’ approach to collecting data is phasing out. * There are now more robust ways to collect qualitative data, these data are a key component of driving innovation. * Important to frame data collection around future use of data; consider more broadly what the data will be used for (e.g. inform policy) and ensure data collection methods can gather this information as directly as possible. | |
| **High value healthcare** | * Common frame of reference/metric for the key components of quality healthcare. * Key components include providers and patients; their respective experiences can be found in qualitative data that can help clarify stakeholder attitudes and behaviours e.g. is there a human error element affecting efficiency or effectiveness or programs, are providers unhappy and unlikely to continue with the program long-term, etc. * Better health outcomes of patients are becoming a larger focus of government regulatory authorities. * A relative new component to high value healthcare is equity; this relates to culturally appropriate care and diversity, and population health. | |
| **Wellbeing** | * There has been a pivot to target wellbeing in the public sector, as support schemes are not independent of well-being type outcomes. This is likely driven by a change in attitudes towards mental health, and acknowledgement that personal life inevitably affects work life. * Larger focus of compensation schemes on social responsibility and improving public health. | |
| **Costing of evaluations** | * No known rubric/matrix for evaluation costing, but anecdotally, organisations often allocate a significant amount of funding into implementation of health programs, but little/not enough into evaluation. * Many factors influencing evaluation pricing, and there is much variation in health programs, e.g. smaller programs may cost more to implement but may be less expensive to scale up. * A key factor of evaluation costs is the ease and availability of capturing data. * Some forms of collecting data are time intensive such as focus groups or interviews, and are more difficult to capture. While these types of data are important, it is more difficult to capture, more expensive, and therefore less commonly collected. | |
| **DVA observations** | * DVA appears to have a useful closed system of providers and consumers. * This alongside high quality, longitudinal utilisation and costing data places them in a strong position for new data analytics e.g. dynamic simulation modelling or forecasting for burden of ill health. | |

Table 23: Case Study III

|  |  |  |
| --- | --- | --- |
| **Australian Government Department of Finance (now Australian Centre for Evaluation)** | | |
| **Interview date and time** | | Tuesday 07 June, 2:00pm – 2:30pm AEST |
| **Theme** | **Key Findings** | |
| **Principles of best practice evaluation** | * Better practice approaches typically have a strategy around evaluations to appropriately prioritise and plan evaluations and make sure evaluation activities are proportional. * Evaluation activity must be robust, ethical and culturally appropriate – particularly with health programs where there may be sensitivities. * Each program should have a monitoring and evaluation framework, this will allow the evaluator to take a risk-based approach to understanding which programs should be evaluated and how they should be evaluated. * Understanding where the risks lie in terms of programs will help to prioritise which programs need to be evaluated. * It is important to understand existing administrative data, as this can be leveraged to support evaluations to drive continuous improvement. A data development agenda may also be needed to guide future collection of data so that it is available when needed. * Understand how evaluation findings can influence government decision-making and how this will impact internal decision making processes and use of resources. Prioritise and schedule evaluation activities to take account of the Budget and Cabinet timeframes to maximise opportunities for findings to be communicated in timely and influential ways. * Feedback loop is critical – evaluation findings need to be used in an effective way. | |
| **Size and value of evaluations for different programs** | * Some departments use a ‘rule of thumb’ approach to funding evaluations (for example, 5%), however, the Department of Finance (now Australian Centre for Evaluation) believes that evaluations need to be fit-for-purpose and a range of factors are relevant to influencing the cost of an evaluation. * Availability of data are a factor in the cost of an evaluation – some programs have administrative systems in place that collect a significant amount of data already, which needs to be accounted for in the cost of the evaluation. * Different programs differ in cost to run, and the level of risk also differs – a rule of thumb approach may not be sufficient for programs that are inexpensive to run but a high risk profile – these evaluations may be more expensive proportionally as they need to investigate whether the program is delivering what is expected. * Evaluation is one aspect of a broader set of assurance mechanisms. | |
| **Type of data to use** | * A mixed methods approach to data is ideal. * Quantitative data from administrative systems are important but do not tell the whole picture. * Qualitative lived experience is equally valid as quantitative data. * The content and context of data are equally important – for example, the environmental context of programs delivered during COVID-19 needs to be considered. * Qualitative data enables the evaluator to tell a powerful story which makes an evaluation more influential – one of the important goals of evaluations is to influence decision making. | |
| **Types of evaluations** | * The type of evaluation to run is dependent on where a program is in its program lifecycle. * As most DVA programs have been in place for a long time, impact evaluations would be more relevant, or, DVA could potentially revisit some formative evaluations to see if there were any issues during implementation of programs. * Important to understand the aim of the evaluation – need to ensure that you are conducting a program level evaluation to understand population level outcomes – this has been a challenge of previous evaluations where evaluators expect the evaluation to answer high level questions, when the program is only a part of that. | |
| **Time to see evaluation impacts** | * 6 months into a program is the first ‘early attention point’ where there are indicators whether an investment in a particular area is on track to deliver the desired benefits – it is important to be realistic about the extent that this can show, as mental health programs, for example, can take multiple years before benefits can be fully visible. * There is an opportunity to leverage learnings from across jurisdictions, including both nationally and internationally to understand benchmarks for timeframes for different programs. | |
| **Change in evaluation practices over time** | * The APS Review found evaluation practices over the last few decades have declined. * The Review indicated that at a Commonwealth level, few organisations had retained evaluation expertise and a lot of external providers have been used to fill skills gaps in evaluation. * The Commonwealth is at the beginning of a long change journey to rebuild evaluation capability across the APS. * There is a move from past approaches where evaluations were seen as a ‘tick and flick’ exercise towards a continuous improvement approach. * Evaluations need to be seen as a constructive part of organisational learning processes – a culture change needs to happen to help people understand that evaluations are an integral part of performance management and effective program delivery. | |