

Science, policy, regulation and audit review checklist John O'Donnell 18 October 2026

Primary audience

- Auditor-Generals
- Productivity Commissions
- Parliamentary inquiries
- Regulatory reviews
- Government departments
- Courts and tribunals
- Scientific review panels
- Large corporations
- Internal auditors

Purpose

This checklist provides a formal science checking, due diligence and audit framework for governments, Auditor-Generals, Productivity Commissions, regulators, parliamentary inquiries, scientific review panels and large organisations.

It assists reviewers to assess whether scientific advice, policies, regulations, programs and management decisions are evidence-based, transparent, accountable, reproducible, proportionate and capable of delivering the intended outcomes.

The checklist applies the principles of scientific due diligence by examining evidence quality, assumptions, methodology, forecasting, implementation, risk, consequences, accountability and post-implementation performance.

The checklist may support:

- scientific audits
- policy reviews
- regulatory reviews
- government inquiries
- research evaluations
- performance audits
- post-implementation reviews
- accountability assessments
- science consequence reviews
- recommendations for corrective or further action.

Review information

Title:

Organisation:

Author(s):

Date:

Reviewer:

Review Scope:

Key matters being assessed

1. Evidence hierarchy assessment

Assessment Question	Evidence	Rating
Is the recommendation based on observation, experiment, modelling or expert opinion?		
What is the highest quality evidence available?		
Were systematic reviews examined?		
Does stronger evidence exist?		

2. Scientific evidence

Assessment Question	Evidence	Rating
Is the scientific basis clearly documented?		
Are original evidence sources available?		
Is evidence current?		
Is evidence representative?		
Is evidence relevant to the decision?		
Are multiple evidence sources available?		
Has causation been distinguished from correlation?		
Are conclusions proportional to evidence?		

3. Scientific methodology

Assessment Question	Evidence	Rating
Is methodology clearly documented?		
Can findings be replicated?		
Has independent review occurred?		
Have competing explanations been examined?		
Are limitations documented?		
Has negative evidence been considered?		

4. Assumptions

Assessment Question	Evidence	Rating
Are assumptions identified?		
Are assumptions tested?		
Are assumptions evidence-based?		
Are implicit assumptions disclosed?		
Have alternative assumptions been assessed?		
Could assumptions fail under changing conditions?		
What evidence would invalidate assumptions?		

5. Reproducibility assessment

Assessment Question	Evidence	Rating
Are datasets accessible?		

Can analyses be independently reproduced?		
Are methods sufficiently transparent?		
Would another reviewer reach the same conclusion?		

6. Uncertainty and knowledge gaps

Assessment Question	Evidence	Rating
Are uncertainties disclosed?		
Are confidence levels stated?		
Are uncertainty ranges quantified?		
Are knowledge gaps identified?		
Could uncertainty alter decisions?		
Have unknown factors been considered?		
Is additional research likely to reduce uncertainty?		

7. Modelling and prediction

Assessment Question	Evidence	Rating
Has modelling been independently reviewed?		
Has modelling been validated against observations?		
Are model limitations disclosed?		
Were sensitivity analyses undertaken?		
Could small assumption changes alter results?		
Have alternative models been considered?		
Are prediction errors reported?		

8. Local reality check

Assessment Question	Evidence	Rating
Do observations support conclusions?		
Do practitioners report similar findings?		
Do historical records support conclusions?		
Are local conditions represented?		
Have examples of failure been assessed?		
Have examples of success been assessed?		

9. Governance and accountability

Assessment Question	Evidence	Rating
Who approved the decision?		
Who is accountable for scientific accuracy?		
Who is accountable for on the ground performance?		
Are responsibilities documented?		
Are review dates specified?		
Are decision records maintained?		
Is independent oversight provided?		

Are outcomes publicly reported?		
Are responsibilities assigned?		
Is independent review proposed?		
Is public reporting required?		
Are review dates specified?		
Is there a process for correction?		

10. Prediction accountability scorecard

Assessment Question	Evidence	Rating
What were the original predictions?		
Were predictions measurable?		
Were predictions achieved?		
If not achieved, why not?		
Were failures reported?		
Were lessons incorporated?		
Has predictive performance improved over time?		
What percentage of predictions proved accurate?		
Are prediction errors publicly reported?		
Is predictive performance improving over time?		
Are poor predictions investigated?		

11. Policy-performance assessment

Assessment Question	Evidence	Rating
Did the policy achieve stated objectives?		
Were outcomes measured?		
Were unintended consequences identified?		
Was independent evaluation undertaken?		
What would likely have occurred if a different approach had been adopted?		
Has a counterfactual been assessed?		
Could observed outcomes have occurred with/ without policy intervention?		
Have current policies ever been tested against the alternatives?		
Were direct and indirect costs quantified?		
Were long term liabilities assessed?		
Were opportunity costs assessed?		
Was value for money demonstrated?		

12. Regulatory burden

Assessment Question	Evidence	Rating
Does compliance increase?		
Does reporting increase?		
Does approval complexity increase?		
Is regulatory duplication occurring?		
Could approval timeframes increase?		
Have less burdensome alternatives been considered?		

13. Cumulative regulation and impact assessment

Assessment Question	Evidence	Rating
Have cumulative compliance burdens been calculated?		
Does this duplicate existing requirements?		
Does the proposal interact with other regulations?		
Have cumulative costs been quantified?		

14. National interest

Assessment Question	Evidence	Rating
Is the proposal in the national interest?		
Does it support sovereign capability?		
Could imports increase?		
Could exports decline?		
Could sovereign risk increase?		
Could strategic national capability decline?		

15. Sovereign capability assessment

Assessment Question	Evidence	Rating
Does the proposal increase reliance on imports?		
Does it affect domestic skills development?		
Could capability become concentrated offshore?		
Could strategic supply chains be weakened?		

16. Consequences of error

Assessment Question	Evidence	Rating
What is the economic cost of being wrong?		
What is the safety cost of being wrong?		
What is the environmental cost of being wrong?		
What is the social cost of being wrong?		
Could impacts be irreversible?		
Does evidence justify the risk?		
Who bears the consequences?		

17. Monitoring and adaptive management

Assessment Question	Evidence	Rating
Are objectives clearly defined?		
Are measurable indicators specified?		
Is monitoring in place?		
Is monitoring independent?		
Can management adapt to new evidence?		
Are corrective actions defined?		
Are review triggers specified?		

18. Bias and conflicts of interest

Assessment Question	Evidence	Rating
Are funding sources disclosed?		
Could funding influence outcomes?		
Are institutional conflicts disclosed?		
Have dissenting views been considered?		
Are reviewers independent?		
Is confirmation bias evident?		
Were contrary findings considered equally?		

19. Scientific integrity red flags

Tick all that apply.

- ☐ No baseline data exists
- ☐ No monitoring program exists
- ☐ Reliance on a single study or expert
- ☐ Contradictory evidence ignored
- ☐ Alternative explanations dismissed
- ☐ Success criteria undefined
- ☐ Critics excluded from review
- ☐ Program continues despite poor outcomes
- ☐ Confidence exceeds available evidence
- ☐ Political or commercial pressure evident
- ☐ Uncertainty not disclosed
- ☐ Evidence selectively reported
- ☐ Baselines changed without explanation

- ☐ Monitoring ceased after poor results
- ☐ Benefits claimed without measurement
- ☐ Models treated as established fact
- ☐ Independent review discouraged
- ☐ Predictions never tested
- ☐ No learning mechanism exists
- ☐ Correlation presented as causation
- ☐ No counterfactual considered
- ☐ Economic impacts omitted
- ☐ Local evidence ignored
- ☐ Success criteria altered after implementation
- ☐ Independent replication unavailable
- ☐ Forecast accuracy never assessed
- ☐ Negative outcomes unpublished
- ☐ Monitoring controlled solely by proponents
- ☐ Program expanded before evaluation

20. Auditor-General readiness test

Assessment Question	Evidence	Rating
Is accountability clear?		
Can decisions be traced to evidence?		
Have assumptions been documented?		
Have uncertainties been disclosed?		
Are monitoring arrangements adequate?		
Are correction mechanisms established?		
Can outcomes be independently verified?		

21. Historical baseline assessment

Assessment Question	Evidence	Rating
Were long term historical records reviewed?		
Were historical management systems assessed?		
Were historical outcomes compared?		
Could historical lessons challenge current assumptions?		

22. Management effectiveness assessment

Assessment Question	Evidence	Rating
Could recommendations actually be implemented?		
Was workforce capability adequate?		
Was funding adequate?		
Were operational constraints recognised?		
Could implementation failure explain poor outcomes?		
Are the proposed outcomes realistically achievable given available resources?		

23. Safety, duty of care and disaster assessment

Assessment Question	Evidence	Rating
Were foreseeable safety risks identified?		
Are safety policies adequate?		
Are all fatalities and major injuries adequately investigated and lessons shared?		
Could failure to act increase risk?		
Is there a duty of care policy in place?		
Who is owed a duty of care?		
Were vulnerable communities identified?		
Was mitigation at sound and adequate levels?		
Was community protection at sound and adequate levels?		
Was emergency responder protection at sound and adequate levels?		
Are house and building insurance costs increasing where low levels of mitigation occur?		
In relation to disasters, what funding % is spent on mitigation as opposed to post disaster funding?		
In terms of bushfire disasters, what % of forests receive planned burns per year?		

24. Economics and productivity assessment

Assessment Question	Evidence	Rating
Were economic impacts assessed?		
Were productivity impacts assessed?		
Were regional impacts assessed?		
Were costs of inaction assessed?		

25. Lesson capture and sound listening

Assessment Question	Evidence	Rating
Are learnings and lessons adequately captured in all situations?		
Are captured learnings and lessons adequately complied with in all situations?		

Is there a willingness to share learning opportunities and gained learnings?		
Is there a willingness to listen to all relevant points of view and appropriately consider key input?		

26. Science risk register

Simple table is suggested:

| Risk | Consequence | Likelihood | Severity | Mitigation | Residual Risk |

This would become the bridge between science review and decision making.

Formal review outcome

Scientific confidence rating:

- ☐ High
- ☐ Moderate
- ☐ Low

Evidence quality rating:

- ☐ High
- ☐ Moderate
- ☐ Low

Risk if incorrect:

- ☐ High
- ☐ Moderate
- ☐ Low

Recommendation:

- ☐ Support
- ☐ Support with conditions
- ☐ Further investigation required
- ☐ Not supported
- ☐ Other action/s