Impact of the "Linked Evidence Approach" Method on Policies to Publicly Fund Diagnostic, Staging, and Screening Medical Tests

Tracy L. Merlin, PhD, Janet E. Hiller, PhD, Philip Ryan, FAFPHM

Background: *The linked evidence approach (LEA) is used in health technology assessment (HTA) to evaluate the clinical utility of new medical tests in the absence of direct trial evidence.* **Objective:** *To determine whether use of LEA affects decisions to publicly fund medical tests.* **Methods:** *Australian HTAs that evaluated medical tests before and after LEA was mandated (in 2005) were screened for eligibility. Data were extracted and the impact of LEA and other possible clinical predictors (selected a priori) on funding decisions was modelled. Regression diagnostics were performed to estimate model fit, model specification, and to inform model selection. The unit of analysis was per clinical indication for each new test, so analyses were adjusted for clustering.* **Results:** *83 HTAs (for 173 clinical indications) were eligible from the 259 screened. When health policy was compared before and after 2005, there was an 11% reduction in overall positive funding decisions, including a 25% decrease in "interim" (coverage with evidence development)*

funding decisions. The odds of obtaining interim funding reduced by 98% (odds ratio = 0.02, 95% confidence interval = 0.0005, 0.17), but there was no change in the direction of funding decisions (odds ratio = 1.36, 95% confidence interval = 0.62, 3.01). Across both time periods, when LEA was used there was a very strong likelihood that the medical test would not receive interim funding (χ^2 = 12.63, df = 1, P = 0.001). For positive funding decisions, the strongest predictors were whether or not the new test would replace an existing test and whether the available evidence was limited. **Conclusions:** *The use of LEA did not predict the direction of funding decisions. Application of the method did predict that a "coverage with evidence development" decision was unlikely. This suggests that LEA may reduce decision-maker uncertainty.* **Key words:** *diagnostic test approval; evaluation methodology; systematic review; reimbursement mechanisms; decision making; policy.* **(MDM Policy & Practice 2016;1:1–12)**

In the past, evaluations of medical tests for public funding decisions have been largely restricted to assessments of test accuracy or performance with little consideration given to the impact on patients. This can include receiving a false-negative test result—leading to a delay in treatment—or a false-positive result—leading to inappropriate treatment.[1] Test evaluations have been limited due to the lack of primary research, in particular randomized controlled trials assessing the impact of testing on patient health outcomes.[2]

If a test performs poorly in a trial, false-positive and false-negative test results will be reflected in the measured health outcomes of patients. However, as trial evidence of the impact of medical

tests on the health outcomes of patients is often scarce, policy makers are faced with making decisions on access to, and reimbursement of, diagnostic, staging, and screening tests on the basis of incomplete and uncertain information.

To address this lack of evidence, a methodology was published in 2005 that tries to maximize the amount of critical information presented to Australian policy makers.[3,4] This "linked evidence approach" (LEA) involves the narrative linking of systematically acquired evidence assessing each component of a test-treatment pathway. The aim is to predict the likely impact of testing on patient health outcomes.

Using LEA, systematic literature reviews and meta-analyses (where possible) are conducted on existing research to determine the accuracy of the new test relative to appropriate reference standards, the impact of the new test on clinical decision-making, and—where circumstances are appropriate[5]—the impact of likely treatment choices on patient health

outcomes. These data are used in decision analytic models to determine the effectiveness and cost-effectiveness of the new tests, relative to existing tests.

The method was informed by criteria developed by Fryback and Thornbury[6] to assess the efficacy of diagnostic imaging tests. The method also built upon the analytic frameworks pioneered by the US Preventive Services Task Force and used in clinical practice guideline development.[7] These frameworks address both the harms and benefits of medical testing on the patient.[8–10]

LEA is similar to the US Preventive Services Task Force analytic frameworks in that a systematic review of the evidence supporting each element (key question) in the test-treatment pathway is undertaken. It differs in that there is a comparison of two test-treatment pathways—the *likely* clinical pathway for patients should the new test become available/publicly funded, and the pathway for patients without the new test being available (current practice). The PICO criteria used for determining study eligibility are derived by comparing the two pathways at each linkage stage: 1) to determine comparative test accuracy (relative to an appropriate reference standard for both the new test and comparator test), 2) comparative impact on clinical decision-making/treatment options, and 3) comparative treatment effectiveness. In some cases an "abridged" LEA can be undertaken; this involves a search for evidence on the comparative accuracy of the new medical test and its impact on

clinical decision making, but an evaluation of the comparative effectiveness of the consequent treatment options is not conducted. This latter element would be considered unnecessary if, for example, the test accuracy evidence demonstrated that the *new* medical test identified patients with a similar spectrum of disease to patients currently receiving standard treatment after diagnosis with current tests. Thus, the decision to proceed with the third linkage depends on the findings from the evidence collated earlier in the pathway.[5]

Although LEA was used sporadically in the assessment of medical tests in Australia from 1999, in 2005 the approach was mandated by the federal government[3] when commissioning health technology assessments (HTAs) of medical tests to inform public funding policy decisions. The method is iterative in that a search for direct (trial) evidence is conducted first, and then if the evidence base is insufficient to address the policy question, LEA is undertaken.

The objective of this study was to determine what effect (if any) the use of LEA methodology and other evidentiary factors had on Australian policy makers' decisions to publicly fund diagnostic, staging, and screening tests.

METHODS

The independent committee in Australia that makes decisions regarding the funding of medical tests, through the Medicare Benefits Schedule, is the Medical Services Advisory Committee (MSAC). This committee began making recommendations on the reimbursement of new health technologies in 1999. This study covered the period of HTA production and decision making from 1999 until December 2014, as the format of commissioned HTA reports changed in 2015/2016. Guidance on the assessment of diagnostic technologies using LEA was introduced in August 2005,[3] but as various drafts were produced prior to this release date, the whole of 2005 was considered a changeover period in some of the analyses that have been conducted. Policies and practice with regard to the HTA of medical tests were compared before and after 2005.

HTA reports were included in this study if they met the following criteria:

- Considered by MSAC between 1999 and December 2014

- The whole assessment report was publicly available on the MSAC website (www.msac.gov.au) at the cutoff date of 30 December 2014
- The evaluation was a "contracted assessment" commissioned by the Australian Government Department of Health, irrespective of whether the health technology was identified through an internal referral, an external application for public funding, or was an update of an application previously considered by MSAC
- The report concerned the assessment of a diagnostic, screening, or staging test; definitions of these types of tests have been reported previously[5]

HTAs were excluded from consideration if

- The test being assessed was used to monitor response to therapy
- The test being assessed was pharmacogenetic—as the use of LEA for these tests has been reported elsewhere[11,12]
- The HTA was commercial in confidence, withdrawn, or not produced

Independent duplicate selection and data extraction occurred for 59 of the 173 (approximately one third) test clinical indications that were eligible. The unit of analysis was test evaluation *per clinical indication*, as tests were often used for multiple purposes and thus several evaluations may have been included in one HTA report. Information was extracted from public summary documents regarding the final MSAC funding decision for each medical test.

Data were analyzed using Microsoft Excel 2013 and Stata Version 13. Logistic regression analyses, with robust variance estimation to account for the nonindependence of clustered data, was performed to determine whether use of LEA, or other factors apparent from the evidence base, predicted a decision to reject or support public funding of the test. Clustered variances were likely as the same test was often used for multiple clinical indications and so evaluation methodologies were likely to be similar in each report.

Independent variables selected a priori as possible predictors included the following:

- *Test purpose*: Whether the new test was to be used as an add-on test, replacement test, or triage test[13]
- *Year of decision*: Year that public funding decision was made by MSAC
- *Methodological approach*:
  - "Direct evidence only"—Reporting only on clinical trials assessing the impact of a test on patient health outcomes

  - "Direct evidence plus full LEA"—Reporting on direct evidence and supplementing this with a linkage of evidence on the accuracy of the medical test, its impact on clinical decision making (e.g., changes in patient management), and the effectiveness of consequent treatment options
  - "Direct evidence plus LEA but full linkage not required"—Reporting on direct evidence and supplementing this with an abridged LEA, whereby evidence on the effectiveness of the treatments is not required
  - "Components of LEA"—Reporting on isolated elements of the test-treatment pathway (most commonly, test accuracy alone) with no rationale given for selecting only those elements
  - "Direct evidence plus components of LEA"—Reporting on direct evidence and supplementing this with elements of the test-treatment pathway (most commonly, test accuracy alone) with no rationale given for selecting individual elements
- *Quality of the evidence base*:
  - Poor/not poor quality—Methodological flaws in the evidence base
  - Limited/not limited data—Presence or absence of evidence to address the question(s) (whether for direct evidence or each of the evidence linkages)
  - Low/high applicability—Ability to translate the findings from the evidence base to the proposed population and delivery of care in the local health system
  - Heterogeneity/homogeneity of findings—Degree of consistency in the findings reported from the evidence base
  - Imperfect/accurate reference standard test

The dependent variable was a positive funding decision. There were five types of funding decision—funding supported, funding rejected, interim funding (approximately 5 years of funding before the decision is reviewed or new evidence is presented), keep current funding (after a funding decision is reviewed favorably), or no decision required (these generally occurred when MSAC was asked for an evaluation but the funding decision was made at a jurisdictional level). However, as a positive MSAC funding decision can mean a *new* positive funding decision, the maintenance of existing funding (i.e., an interim funded test that is being reviewed) or the decision to provide interim funding, this dependent variable was a composite.

Various logistic regression models were tested to determine whether LEA and/or the other prespecified independent variables predicted positive funding decisions overall, new funding decisions alone, or interim funding decisions alone. Regression diagnostics were conducted to confirm model specification and to determine model fit. The Wald statistical test was used to test the hypothesis that the maximum likelihood estimate of the parameters of interest in each model predicted the proposed true value.[14] Model selection was primarily informed by Akaike information criterion measures to estimate minimization of information loss.[15]

There was no external funding source for this study. The authors performed the research independently and as part of their role as academics at the University of Adelaide.

RESULTS

Of the 259 HTAs available on the MSAC website, 83 were found to meet the eligibility criteria and reported on the use of a test for diagnosis (61%), staging (23%), or screening (12%) purposes for 173 clinical indications. Nearly one half of these were "add on" tests (42%), while approximately one quarter (26%) were "replacement" tests.

Thirty-nine evaluations of diagnostic, staging, or screening tests conducted before LEA was introduced (May 1999 to August 2005), and comprising 63 clinical indications, were compared to 44 evaluations of tests (110 clinical indications) conducted after LEA was introduced (August 2005 to December 2014).

HTA Methodology

A comparison of evaluation methodologies before and after 2005 indicates that use of the "components of LEA" reduced significantly (Figure 1). "Components of LEA" predominantly only considers diagnostic accuracy data and not the downstream effects of a test. Between 2005 and 2010, the use of "components of LEA" ceased completely, only to reemerge—albeit to a lesser extent—between 2011 and 2014.

MSAC Funding Decisions

Before the introduction of the MSAC guidelines for evaluating diagnostic tests (i.e., between May 1999 and July 2005), 63 clinical indications for eligible diagnostic, staging, or screening tests were assessed to determine if there was sufficient evidence of test safety, effectiveness, and cost-effectiveness to warrant public funding through the Medicare Benefits Scheme. Sixty-four percent of the funding decisions were positive. This included 27% where the decision was conditional upon a review in 5 years (interim funding) and 5% where the original interim funding decision was confirmed after review. Thirty-five percent of the funding decisions were negative.

After the LEA methodology was formally introduced (between August 2005 and December 2014), 110 specific uses of diagnostic, staging, or screening tests were evaluated for a public funding decision. Fifty-nine percent of these funding decisions proved to be positive. This included 1% where the decision was conditional upon a review in 5 years (interim funding) and 11% where the original interim funding decision was confirmed after review. Thirty-eight percent of the funding decisions were negative.

The most common methodological approach in the reports supporting these decisions was a search for direct evidence, supplemented by an LEA. In most cases there was limited direct evidence available, or the available evidence concerned a population or intervention that was not perfectly applicable, and so a full LEA was undertaken to redress shortfalls in the evidence base.

A comparison of funding decisions before and after the use of LEA was recommended indicates that the proportion of funding decisions informed by the method increased substantially (Figure 2). The odds of an ensuing negative funding decision was five times higher than for a positive funding decision with HTAs that only used "components of LEA" during the period 2005 to 2014, although there was significant uncertainty about the estimate (unadjusted odds ratio [OR] 5.37; 95% confidence interval [CI] 0.50, 269.41).

Predicting Test Funding

The models with the best fit at predicting an overall positive funding decision are depicted in Table 1, but the prediction capabilities of three of these four models were still consistent with chance.

Model 1 predicted that a public funding decision was dependent on two factors—the absence/presence of limited data and the use of the new test as a replacement for an existing test ($\chi^2 = 6.63$, df = 2, $P = 0.04$). When Models 1 and 2 were compared, the
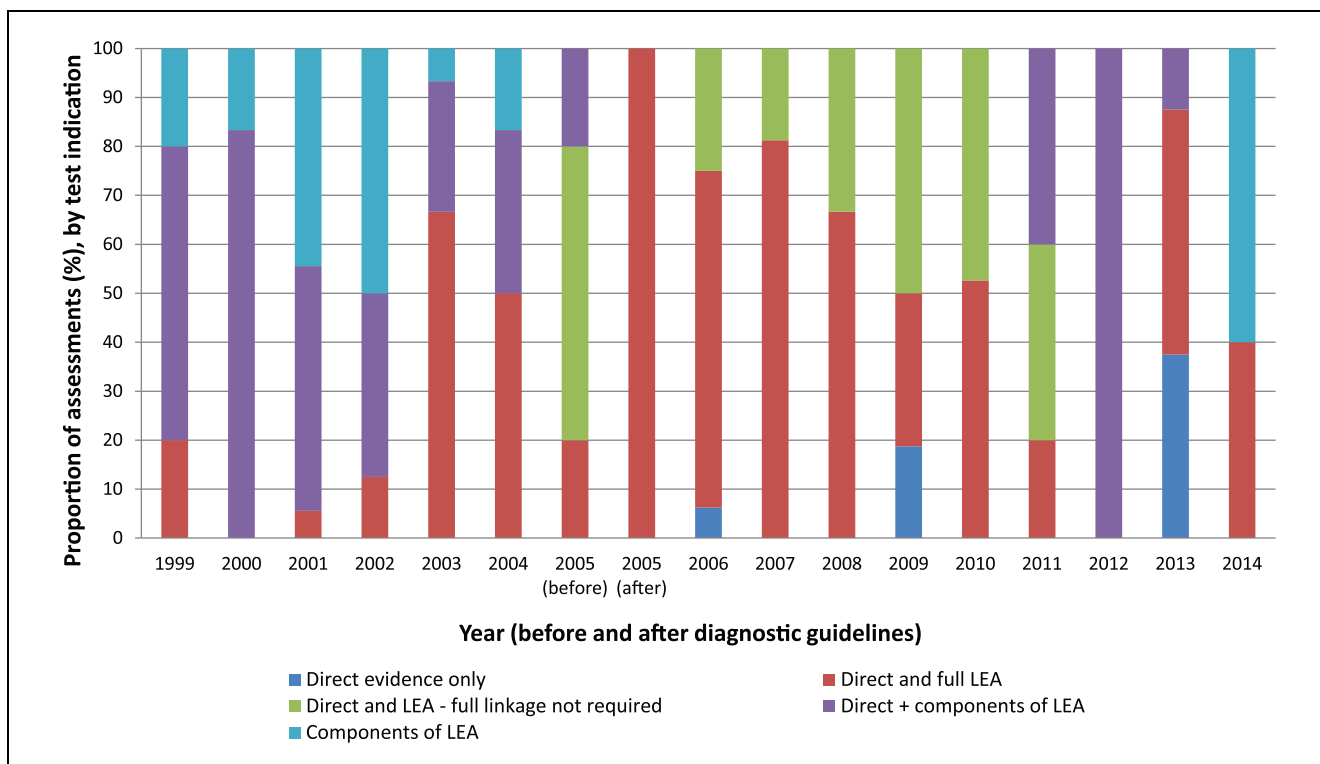
Figure 1   Change in evaluation methodology over time.
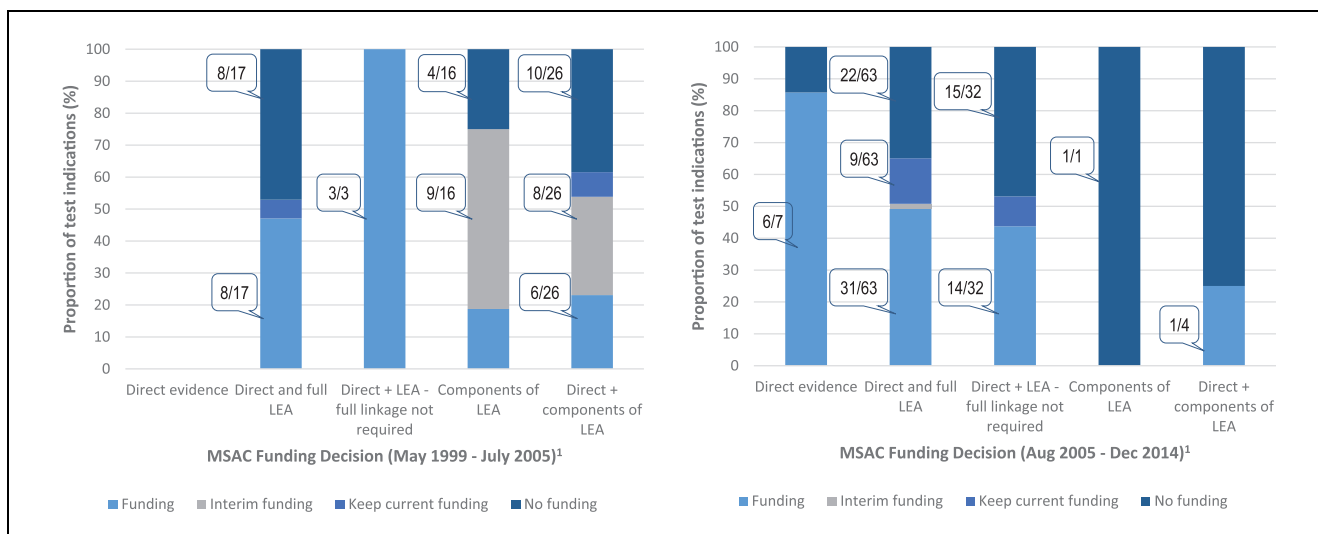
Figure 2   Funding decisions by methodological approach.
[1]*Where no funding decision was made, the HTA has been excluded.*

difference was not marked but Model 1 had a slightly better fit to the data.

Other combinations of the prespecified independent variables, including methodological approach (specifically, the use or not of LEA), were not significant predictors (data not shown). There was no apparent association between decision year (1999–2014) and public funding decisions, nor was there

Table 1 Predicting Overall Funding of Medical Tests in Australia

Variable	Model 1 β [SE]	Model 1 Robust OR$_{adj}$ [95% CI]	Model 2 β [SE]	Model 2 Robust OR$_{adj}$ [95% CI]	Model 3 β [SE]	Model 3 Robust OR$_{adj}$ [95% CI]	Model 4 β [SE]	Model 4 Robust OR$_{adj}$ [95% CI]
Constant	0.88 [0.41]	2.42 [1.08, 5.39]	1.02 [0.48]	2.76 [1.08, 7.02]	0.81 [0.53]	2.25 [0.80, 6.31]	1.04 [0.48]	2.82 [1.10, 7.28]
Test purpose								
"Replacement" v. "not replacement" for existing test	0.79 [0.43]	2.21 [0.94, 5.18]	0.78 [0.44]	2.17 [0.91, 5.17]	0.87 [0.45]	2.39 [0.98, 5.81]	0.83 [0.45]	2.30 [0.96, 5.50]
Reference standard								
Accurate v. imperfect			−0.29 [0.41]	0.75 [0.33, 1.68]				
Evidence quality								
Poor v. good quality							−0.21 [0.41]	0.81 [0.36, 1.82]
"Limited" v. "not limited" data available	−0.78 [0.42]	0.46 [0.20, 1.05]	−0.76 [0.43]	0.47 [0.20, 1.09]	−0.92 [0.46]	**0.40 [0.16, 0.98]**	−0.82 [0.44]	0.44 [0.19, 1.03]
LEA v. "no LEA" methodology					0.19 [0.52]	1.21 [0.44, 3.35]		
No. of observations	169, adjusted for 81 clusters		169, adjusted for 81 clusters		162, adjusted for 77 clusters		169, adjusted for 81 clusters	
Log pseudo-likelihood	−107.79		−107.39		−103.46		−107.60	
Wald test	**χ² = 6.63, df = 2, P = 0.04**		χ² = 7.61, df = 3, P = 0.06		χ² = 6.97, df = 3, P = 0.07		χ² = 6.53, df = 3, P = 0.09	
Pseudo R^2	4%		4%		4%		4%	
AIC	1.31		1.32		1.33		1.32	
AIC * n	221.59		222.78		214.93		223.20	

Note: Bold indicates statistically significant predictor. OR$_{adj}$ = odds ratio adjusted for other predictors in the model; CI = confidence interval; β = beta coefficient; SE = standard error; LEA = linked evidence approach; AIC = Akaike information criterion.

an apparent difference in decisions between time periods, that is, before or after introduction of LEA. Results were similar irrespective of whether the year of introducing LEA (2005) was included when time periods were compared (data not shown).

With regard to test purpose, both add-on tests and replacement tests predicted public funding decisions (add on tests: unadjusted OR 2.8, 95% CI 0.97, 8.10, P = 0.06; replacement tests: unadjusted OR 4.66, 95% CI 1.40, 15.57, P = 0.01), although replacement tests were the stronger predictor. This is not surprising as replacement tests are more likely to be cost-effective or cost-saving than add-on tests. Tests undertaken to triage patients were not associated with a particular type of public funding decision (unadjusted OR 2.0, 95% CI 0.59, 6.84, P = 0.27) but this result was based on only 32 triage tests out of the 173 tests considered.

Predicting New Test Funding

Following the introduction of LEA, positive funding decisions reduced overall by 11%; driven by a 25% reduction in time-limited "interim" funding decisions. "Definitive" positive funding decisions *increased* by 15% but the ratio to negative funding decisions was not significantly different between the time periods (unadjusted OR = 1.36, 95% CI 0.62, 3.01, χ^2 = 0.69, P = 0.41).

Four of the better models at predicting *new* public funding decisions are given in Table 2.

Two of these models (Models 1 and 2) demonstrated a greater than chance ability to predict funding of new tests. Both models were driven by the association between decision making and the presence/absence of limited data. Funding was also affected by whether the new test was a replacement for an existing test (Model 1: χ^2 = 7.22, df = 2, P = 0.03). Model 2 also incorporated an appropriate reference standard as a prediction variable (χ^2 = 8.88, df = 3, P = 0.03). In a comparison between Model 1 and 2, Model 2 had a slightly better fit to the data but the difference was weak. Other possible combinations of the independent variables, including methodological approach, did not predict new

Table 2 Disaggregated Results: Predicting New or Interim Funding of Medical Tests in Australia

| | Predicting *New Funding* | | | | | | | | Predicting *Interim Funding* (5-Year Time-Limited) | | | | | | | |
| | Model 1 | | Model 2 | | Model 3 | | Model 4 | | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
	β [SE]	Robust OR$_{adj}$ [95% CI]	β [SE]	Robust OR$_{adj}$ [95% CI]	β [SE]	Robust OR$_{adj}$ [95% CI]	β [SE]	Robust OR$_{adj}$ [95% CI]	β [SE]	Robust OR$_{adj}$ [95% CI]	β [SE]	Robust OR$_{adj}$ [95% CI]	β [SE]	Robust OR$_{adj}$ [95% CI]	β [SE]	Robust OR$_{adj}$ [95% CI]
Constant	0.84 [0.41]	2.32 [1.03, 5.23]	1.02 [0.48]	2.81 [1.07, 7.35]	0.95 [0.56]	2.57 [0.86, 7.75]	0.89 [0.48]	2.43 [0.94, 6.26]	−0.66 [0.52]	0.52 [0.18, 1.44]	−0.32 [0.61]	0.73 [0.22, 2.42]	−0.06 [0.79]	0.94 [0.20, 4.39]	−0.92 [0.81]	0.40 [0.08, 1.96]
Test purpose																
"Replacement" v. "not replacement" for existing test	0.72 [0.45]	2.06 [0.86, 4.97]	0.78 [0.44]	2.08 [0.85, 5.07]	0.82 [0.48]	2.27 [0.89, 5.77]	0.73 [0.45]	2.08 [0.86, 5.05]								
Reference standard																
Accurate v. imperfect			−0.29 [0.41]	0.66 [0.29 1.53]	−0.47 [0.45]	0.63 [0.26, 1.53]									−1.22 [0.66]	0.30 [0.08, 1.08]
Evidence quality																
Poor v. good quality							−0.06 [0.42]	0.94 [0.41, 2.15]					−0.32 [0.58]	0.72 [0.23, 2.26]	−0.02 [0.50]	0.98 [0.37, 2.59]
"Limited" v. "not limited" data available	−0.92 [0.43]	**0.40** **[0.17, 0.92]**	−0.76 [0.43]	**0.40** **[0.17, 0.94]**	−1.11 [0.47]	**0.33** **[0.13, 0.83]**	−0.93 [0.43]	**0.40** **[0.17, 0.92]**			−0.68 [0.59]	0.51 [0.16, 1.62]	−0.73 [0.61]	0.48 [0.14, 1.60]	−1.30 [0.56]	**0.27** **[0.09, 0.82]**
LEA v. "no LEA" methodology					0.31 [0.57]	1.36 [0.45, 4.16]			−4.08 [1.15]	**0.02** **[0.002, 0.16]**	−3.91 [1.13]	**0.02** **[0.002, 0.18]**	−3.94 [1.16]	**0.02** **[0.002, 0.19]**		
No. of observations	154, adjusted for 75 clusters		154, adjusted for 75 clusters		147, adjusted for 71 clusters		154, adjusted for 75 clusters		166, adjusted for 79 clusters		166, adjusted for 79 clusters		166, adjusted for 79 clusters		173, adjusted for 83 clusters	
Log pseudo-likelihood	−99.90		−99.15		−94.66		−99.88		−37.80		−37.12		−36.97		−51.91	
Wald test	**χ² = 7.22, df = 2, P = 0.03**		**χ² = 8.88, df = 3, P = 0.03**		χ² = 9.16, df = 4, P = 0.06		χ² = 7.16, df = 3, P = 0.07		**χ² = 12.63, df = 1, P = 0.001**		**χ² = 12.74, df = 2, P = 0.001**		**χ² = 13.33, df = 3, P = 0.001**		**χ² = 8.53, df = 3, P = 0.04**	
Pseudo *R²*	4%		5%		6%		4%		34%		35%		35%		10%	
AIC	1.34		1.32		1.36		1.35		0.48		0.48		0.49		0.65	
AIC * n	205.80		222.78		199.32		207.77		79.60		80.25		81.93		111.81	

Note: Bold indicates statistically significant predictor. OR$_{adj}$ = odds ratio adjusted for other predictors in the model; CI = confidence interval; β = beta coefficient; SE = standard error; LEA = linked evidence approach; AIC = Akaike information criterion.

funding decisions greater than chance (data not shown). The use of LEA did not appear to predict a new positive funding decision.

Interim Funding Decisions

After LEA was mandated, the odds of *interim* funding reduced by 98% (unadjusted OR = 0.02, 95% CI 0.0005, 0.17, χ^2 = 26.44, $P < 0.001$).

Four of the better models at predicting interim public funding decisions are given in Table 2. The simplest model (Model 1), with only LEA methodological approach as a predictor, was strongly explanatory of interim funding decisions (χ^2 = 12.63, df = 1, $P < 0.001$). The other three models, with additional independent variables included, were statistically significantly predictive of interim funding, but while the fit of Models 1, 2, and 3 were similar, Model 1 is preferred on the grounds of parsimony.

Models that included LEA were, on the whole, statistically significant because of the strong association between LEA and interim funding decisions—when LEA methodology was used, medical tests did not receive interim funding. There was only one model tested (Model 4) that demonstrated the ability to predict interim funding in the absence of the LEA variable. In this model, if there was poor-quality evidence and limited data available, as well as an imperfect reference standard, then it was likely that interim funding would not be received (χ^2 = 8.53, df = 3, $P = 0.04$).

DISCUSSION

We hypothesized that the mandated use of LEA in Australia would change the way that medical tests were evaluated for their clinical utility. The results of our study have confirmed this. LEA methodology was the most common method for presenting data to policy makers between 2005 and 2014. Commissioned HTA assessors have followed the 2005 *MSAC Guidelines for the Assessment of Diagnostic Technologies*.[3] Decision making was based on the linkage of systematically reviewed evidence on test performance, relative to an accepted reference standard, to evidence on the impact of the test on treatment decisions and through careful consideration of the likely impact of the test on patient health outcomes—including the impact of false-positive and false-negative test results. At the least, it is likely that this could have led to more informed decision making.

After 5 years the presentation of "unlinked" component evidence (primarily the presentation of technical accuracy alone) has reemerged. There are several reasons why this may have occurred. HTA processes in Australia were reviewed in 2009,[16] and reforms of the process led to the introduction of an option for applicants for public funding to submit their own assessments of medical tests, which were then critiqued by independent contracted assessors, to facilitate a potentially faster review by the decision-making body.[17] The option to have the assessment conducted by independent assessors, essentially "free of charge," was still available but timeliness could not be guaranteed.

To mitigate the effects of this change in process, our study has only included contracted assessments not submission-based assessments. But unintended consequences of the change process may also have affected contracted assessments. New guidance needed to be developed to assist both applicants and contracted assessors in their evaluation of medical tests. The 2005 guidance was archived on the MSAC website but the new guidance (which incorporates LEA methodology) and templates for presenting HTAs on investigative technologies were only released in 2016. It is possible that assessors who were commissioned between 2010 and 2015 did not have access to "best practice" guidance on the use of LEA in the analysis of tests. Alternatively, it is possible that the proportion of direct evidence available for these assessments was sufficient so that an explicit linkage of evidence was not needed and test accuracy data were provided only for the sake of completeness. Irrespective of the reason it is apparent that contracted assessments that did not use LEA in recent years tended to report on tests that were subsequently rejected for public funding, although the numbers were too small to establish this as occurring above chance.

The results of the logistic regression models indicate that the choice of methodological approach is unlikely to affect the direction (positive or negative) of funding decisions, but use of LEA is strongly associated with a negative likelihood of a medical test obtaining interim funding.

The additional information provided in a LEA could plausibly reduce the uncertainty associated with decision making and therefore reduce the need to make interim funding decisions. The observed concomitant increase in more definitive positive or negative public funding recommendations might have been the result of changes in the

quantity and coherence (prediction of clinical utility) of information provided to decision-makers.

However, the data used in this study were uncontrolled and so other explanations for the change in policy behavior cannot be ruled out. The reduction in interim funding after 2005 could have been the consequence of a policy change at the government level or due to turnover in the composition of the decision-making committee.

Ideally, the impact of LEA would be tested in a prospective study that compared funding outcomes or recommendations from evidence synthesized using the LEA approach or synthesized using other methods. There would need to be adequate adjustment for potential confounders, such as those independent variables used in our regression analyses. Prospective data collection in both study arms would mitigate any structural or other changes occurring at the decision-making level.

Although a prospective analysis could not be undertaken in the current study, the association between LEA and the reduction in interim funding held for the 87% ($n = 110$ indications) of HTAs that used the method in the period after 2005 as well as in the period *prior* to 2005 for nearly one third of HTAs (32%, $n = 63$ indications) that also used LEA (Figure 2). If LEA is one of the causes for this change in policy behavior, then use of the approach might obviate the need for ''coverage with evidence development'' arrangements for some services that involve medical tests. Coverage with Evidence Development ''is characterized by restricted coverage for a new technology in parallel with targeted research when the stated goal of the research or data collection is to provide definitive evidence for the clinical or cost-effectiveness impact of the new technology.''[18p79] In the case of medical tests, it is probable that uncertainty will be the norm, as direct evidence of the impact of testing on health outcomes is rare[2] and so decision-maker uncertainty is likely to be high. However, if sufficient linked evidence is already available then there is no need to generate new information to reduce that decision-maker uncertainty. The available evidence simply needs to be identified and selected appropriately and used systematically in decision modelling to predict likely health outcomes.

This does not mean that the use of LEA will always result in certainty. LEA is also affected by the availability of information. If a positive finding using the new test results in additional cases being detected, meaning that the spectrum of disease in the currently diagnosed population changes, then evidence will be needed on how existing treatments perform in this broader population. If these data are unavailable, then a linked evidence approach will not be informative,[3,5] and there may be a case for coverage with evidence development, particularly in areas of high unmet clinical need.[18]

CONCLUSION

The use of LEA did not affect the direction of reimbursement decisions to any great extent. Fewer interim funding decisions after introduction of the methodology tends to suggest greater decision-maker certainty regarding the clinical utility of medical tests, although other explanations for this finding cannot be ruled out. Whether the use of LEA in HTA has also resulted in *better* decision making with regard to the funding of medical tests is an issue for future research.

ACKNOWLEDGMENTS

REFERENCES

1. Staub L, Dyer S, Lord S, Simes RJ. Linking the evidence: intermediate outcomes in medical test assessments. Int J Technol Assess Health Care. 2012;28(1):52–8.

2. Ferrante di Ruffano L, Davenport C, Eisinga A, Hyde C, Deeks JJ. A capture-recapture analysis demonstrated that randomized controlled trials evaluating the impact of diagnostic tests on patient outcomes are rare. J Clin Epidemiol. 2012;65(3):282–7.

3. Medical Services Advisory Committee. Guidelines for the Assessment of Diagnostic Technologies. Canberra, ACT: Commonwealth of Australia; August 2005.

4. Lord SJ, Irwig L, Bossuyt PM. Using the principles of randomized controlled trial design to guide test evaluation. Med Decis Making. 2009;29(5):E1–E12.

5. Merlin T, Lehman S, Hiller JE, Ryan P. The ''linked evidence approach'' to assess medical tests: a critical analysis. Int J Technol Assess Health Care. 2013;29(3):343–50.

6. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making. 1991;11(2):88–94.

7. Harris R, Helfand M, Woolf S, et al. Current methods of the US Preventive Services Task Force: a review of the process. Am J Prev Med. 2001;20(3 Suppl.):21–35.

8. Nelson H, Huffman L, Fu R, Harris E. Genetic risk assessment and BRCA mutation testing for breast and ovarian cancer susceptibility: systematic evidence review for the U.S. Preventive Services Task Force. Ann Intern Med. 2005;143(5):362–79.

9. Nelson H, Pappas M, Zakher B, Mitchell J, Okinaka-Hu L, Fu R. Risk assessment, genetic counseling, and genetic testing for BRCA-related cancer in women: a systematic review to update the U.S. Preventive Services Task Force Recommendation. Ann Intern Med. 2014;160(4):255–66.

10. Whitlock E, Garlitz B, Harris E, Bell T, Smith P. Screening for hereditary hemochromatosis: a systematic review for the U.S. Preventive Services Task Force. Ann Intern Med. 2006;145(3):209–23.

11. Merlin T, Farah C, Schubert C, Mitchell A, Hiller J, Ryan P. Assessing personalized medicines in Australia: a national framework for reviewing codependent technologies. Med Decis Making. 2013;33(3):333–42.

12. Merlin T. The use of the "linked evidence approach" to guide policy on the reimbursement of personalized medicines. Personalized Med. 2014;11(4):435–48.

13. Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. BMJ. 2006;332(7549):1089–92.

14. Wald A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans Am Math Soc. 1943;54:426–82.

15. Akaike H. A new look at the statistical model identification IEEE Trans Automat Control. 1974;19(6):716–23.

16. Australian Government, Department of Health and Ageing. Review of Health Technology Assessment in Australia—A Discussion Paper. Canberra, ACT: Commonwealth of Australia; 2009.

17. Medical Services Advisory Committee. Technical guidelines for preparing assessment reports for the Medical Services Advisory Committee - Service Type: Investigative (Version 2.0). Canberra, ACT: Australian Government Department of Health; March 2016.

18. Trueman P, Grainger D, Downs KE. Coverage with evidence development: applications and issues. Int J Technol Assess Health Care. 2010;26(1):79–85.