International Journal of Technology Assessment in Health Care, **27:2** (2011), 159–168.

doi:10.1017/S0266462311000018

POLICIES

Coverage with evidence development: The Ontario experience

Leslie Levin
University of Toronto and *Ministry of Health and Long-Term Care*

Ron Goeree
McMaster University and *St. Joseph's Healthcare Hamilton*

Mark Levine
McMaster University and *Ontario Clinical Oncology Group*

Murray Krahn
University of Toronto and *Toronto Health and Technology Assessment Collaborative*

Tony Easty
University of Toronto and *University Health Network*

Adalstein Brown
University of Toronto and *St. Michael's Hospital*

David Henry
University of Toronto and *Institute for Clinical Evaluative Sciences (ICES)*

Background: For non-drug technologies, there is often residual uncertainty following systematic review, mainly due to inadequate evidence of efficacy. The unwillingness to make decisions in the presence of uncertainty may lead to passive diffusion and intuitive decision making with or without public pressure. This may affect health system sustainability. There is increasing interest in post-market evaluation through processes that include coverage with evidence development (CED) to address residual uncertainty regarding effectiveness and cost-effectiveness. Global experience of CED has been slow to develop despite their potential contribution to decision making.
Methods: Ontario's field evaluation program to better inform decision making represents a collaboration between physicians, policy decision makers and academic centers. We report results of the first ten CEDs from this program to assess whether they achieved their objective of influencing policy by addressing residual uncertainty following systematic review.

The authors thank Ms. Kellee Kaulback for her assistance in the preparation of the manuscript.

Results: Since 2003, nineteen field evaluation studies to resolve residual uncertainty following systematic review have been completed, ten of which met the criteria of CED and are the focus of this report. There was more than one patient subgroup or intervention in three of the CEDs. This provided the basis for evaluating thirteen outcomes. In each case, the CED addressed the uncertainty and led to a decision based on the systematic review and CED result. The CEDs led to adoption of the technology in six instances, modified adoption in three instances and withdrawal in four instances.
Conclusions: CED makes an important contribution to translating evidence to decision making. Methodologies are needed to increase the scope and reduce timelines for CEDs, such as the use of linked comprehensive and robust data sets and collaborative studies with other jurisdictions. CED before making long-term funding decisions, especially where there is uncertainty of effectiveness, safety or cost-effectiveness, should be increasingly funded by health systems.

Keywords: Coverage with evidence development, Post-market evaluation, Health systems, Policy, Evidence-based decisions, Field evaluation studies, Medical Advisory Secretariat, Ontario Health Technology Advisory Committee

Evidence is being increasingly used to assist in making appropriate choices in the adoption of health technologies (3;20). However, the quality of existing evidence is often insufficient to address the needs of decision makers. Evidence may be lacking or may not be generalizable across jurisdictions. For these reasons, contextualized "real-world" effectiveness data may be useful in aiding the decision-making process (28). For most health systems, failure to make a policy decision in the face inadequate evidence may lead to passive diffusion, no diffusion, or intuitive policy decision making.

In 2006, the Center for Medicare and Medicaid in the United States (US) instituted coverage with evidence development (CED) which offered an option for coverage of promising drugs, biologics, devices, diagnostics, and procedures that would not otherwise meet Medicare's evidentiary standards for being "reasonable and necessary" (27). It provides an option to fund a technology to collect primary data that informs decision making when uncertainty exists regarding efficacy or effectiveness. The United Kingdom-based National Institute for Health and Clinical Excellence in 2003 developed the capacity to promote additional research through its Only-In-Research program to gather primary evidence and inform guidance development. Other initiatives gave rise to registries (2;19), but there are few successful examples of clinical trials that were sponsored to inform decision making. One rare example is a randomized controlled clinical trial (RCT) that assessed the efficacy of high dose chemotherapy with bone marrow support in the treatment of breast cancer (23). With an increasing interest in CED, we report on our experience using CED in the province of Ontario where coverage for a technology was provided conditional upon additional data being collected to specifically address residual uncertainty to better inform evidence-based decision making.

In 2003, the Ontario Health Technology Advisory Committee (OHTAC) was created as an arms-length evidence-based advisory committee to the Ministry of Health and Long-Term Care (MOHLTC) for non-drug health technologies. Residual uncertainty following analysis of available evidence by the Medical Advisory Secretariat (MAS) sometimes prevented OHTAC from making a recommendation regarding the adoption of a technology, even though patient benefits were potentially substantial. A field evaluation program to evaluate these technologies to address this uncertainty was subsequently developed by MAS in collaboration with several academic partners (9;30).

Since 2003, the field evaluation program has included a diverse range of health technologies. Nineteen field evaluation studies have been completed and an additional nineteen are under way. Of the nineteen completed studies, ten meet the definition of CED and are the focus of this study. We will specifically focus on the design, conduct, and impact of the CEDs on reducing residual uncertainty, impacting policy, and influencing changes in usage.

Features of the Current Ontario Program

Details of the evidence-based framework that guides the uptake and diffusion of health technologies in Ontario are published elsewhere (9;12;30). Briefly, the evidentiary core that feeds into OHTAC consists of peer-reviewed contextualized 16-week systematic reviews conducted by MAS to assess effectiveness. Synchronous economic analyses are undertaken by the Toronto Health Economic and Technology Assessment Collaborative (THETA) at the University of Toronto and the Program for the Assessment of Technologies in Health Research Institute (PATH) at McMaster University. Field evaluations are conducted by the Ontario Clinical Oncology Group (OCOG), PATH, THETA, and the University Health Network (UHN) Healthcare Human Factors Group. In addition, the Institute for Clinical Evaluative Sciences (ICES) houses linked health services data to support field evaluations.

The MOHLTC provides core support to these institutions and incremental allocations for each field evaluation. The

total cost of each field evaluation is estimated at CAN$600,000 which includes protocol development and implementation and costs attributed to data collection, analysis and reporting. This estimate does not include costs absorbed by institutions or of the technology being tested. The process provides an opportunity to bend the diffusion—and, therefore, the cost-curves—for these technologies.

Most often the recommendation for a field evaluation is based on the quality of evidence of effectiveness assessed as part of the systematic review and determined by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) process (1). Using GRADE criteria, further research is deemed less likely to affect the confidence in the estimate when the quality of evidence is moderate to high. Therefore, low to very low quality evidence often triggers consideration for a field evaluation if the technology is deemed to have potentially important effects. The assessment of cost-effectiveness, social values, and feasibility of implementation are additional components of the OHTAC decision-making process (11) that may also trigger a field evaluation, as may concerns regarding generalizability.

DESIGN OF FIELD EVALUATIONS

The design of each field evaluation varies according to the nature of the residual uncertainty. Randomized controlled trials (RCTs) are considered when there is no moderate to high quality evidence of efficacy; prospective observational trials evaluate "real-world" effectiveness, safety or accuracy; and registry studies measure real-world effectiveness or safety. The utility of registry studies is maximized when patient outcomes data are linked to robust administrative databases.

Decision analytic policy models have been built to examine effectiveness of discrete interventions, within the context of other interventions used for the same condition over variable time horizons. These are based on Ontario patient outcomes, cost, and (usually) epidemiological data and, therefore, provide realistic and grounded estimates of potential costs and outcomes to decision makers. Models built to date are for diabetes, cardiovascular disease, and pressure ulcers, but for the purpose of this article, only experience with the diabetes model will be presented. When using these models to examine the effects of new technologies, probabilities to populate the model are most often derived from the systematic review. It is anticipated that these models will become especially useful in setting long-term strategies and planning for chronic disease management.

RATIONALE, RESULTS, AND POLICY IMPLICATIONS OF COMPLETED CEDS

Table 1 shows the question asked in each CED, the nature of the evaluation undertaken, results, and policy decisions taken. Additionally, a study on deep brain stimulation was discontinued before the study commenced due to difficulties in executing a jointly funded arrangement and a study on negative wound pressure for chronic wound healing was discontinued due to poor patient accruals. The latter study raised concerns regarding difficulties in conducting research to capture primary patient data in a community setting. Two separate patient subgroups were considered in one CED related to the evaluation of endovascular aortic aneurysm repair (EVAAR) in high surgical risk and in low surgical risk populations. Four separate interventions were considered in another CED related to the evaluation of interventions in the management of diabetes, these being multidisciplinary care, behavior modification, insulin infusion pumps in the management of insulin-dependent type 2 diabetes, and bariatric surgery in the treatment of morbidly obese individuals with type 2 diabetes. One positron emission tomography (PET) study is awaiting final analysis. Twelve interventions and one patient subgroup will therefore be presented.

Positron Emission Tomography (PET) – The Use of CED to Address Uncertainty Relating to Clinical Utility

For the analysis of diagnostic technologies, both accuracy and clinical utility are important outcomes. Clinical utility defines how a technology influences clinical decision making and affects patient outcomes. Examples of field evaluations to explore clinical utility for diagnostic technologies are provided by PET scanning, described in this section, and CT angiography, discussed in the section dealing with assessment of generalizability below.

When a decision regarding the adoption of PET scanning was first considered in 2002, few studies addressed clinical utility (10), and yet the potential cost to the health system was considerable. Three alternatives were: (i) to decline investment until evidence of clinical utility was available, (ii) to adopt an open-ended funding strategy with little evidence of clinical utility, or (iii) to evaluate PET scanning and to fund it according to evidence of its clinical utility in different disease conditions. The latter option was chosen, overseen by a provincial expert steering committee, with clinical trials conducted by OCOG. A description of this evaluation process for cancer indications has been published (7). The steering committee also monitors published evidence of clinical utility and makes recommendations for insurability based on these reports while the evaluation program conducts PET studies as recommended by the steering committee.

Five clinical trials have completed accrual to assess the clinical utility of PET scanning in early (13) and locally advanced (29) non–small-cell lung (NSCL) cancers, early stage breast cancer (22), head and neck cancer and colon cancer metastatic to liver (Table 1).

PET in Early Stage NSCL Cancer

Rationale. There was originally conflicting evidence and therefore uncertainty as to whether PET scanning

https://doi.org/10.1017/S0266462311000018 Published online by Cambridge University Press

Table 1. Completed Field Evaluations Meeting the Definition of CED and Policy Development

Technology [*N*]	Overseen BY	Type of study	Reason for field evaluation	Result	Policy decision
Drug eluting stents (DES) [21,000] (26)	PATH, with ICES, Cardiac Care Network and 18 leading cardiologists	Prospective pragmatic registry-based	Testing generalizability of RCT evidence and cost-effective analysis	Only effective in patients at high risk for restenosis	Funded 30% conversion from bare-metal to DES[b] (Compare 90% in USA)
Endovascular abdominal aortic aneurysm repair (EVAR) [160] (25)	PATH and single AHSC	Prospective observational	Safety assessment of endoleak	No endoleak. Cost-effective for high surgical risk patients but not for low-risk	Improve access to EVAAR for high risk[a] but not fund for low surgical risk[c]
Multifaceted primary care diabetes intervention program (21)	PATH, working with Oxford University	Before and after study design using a micro simulation economic model	Model to estimate long-term downstream effects of diabetes interventions help prioritize funding for a provincial diabetes strategy following systemic review of each component of the strategy	Allowed downstream outcomes & costs based on interacting variables to be determined for different time horizons	Cost-effective strategies were bariatric surgery for morbidly obese diabetics[a], and multidisciplinary treatment teams[a]. Insulin infusion pumps for type 2 diabetes not cost-effective and not funded[c]
64-slice CT angiography v coronary angiography (CA) [175] (4)	PATH, working with cardiologists, radiologists, selected academic health science centers	Patients referred for CA also undergo CT angiography	MAS HTA concluded uncertainty regarding system impacts, indications for use, and parameters for diffusion	Sensitivity much lower than reported in RCTs reducing cost-effectiveness	OHTAC recommendation cautious adoption in AHSCs until issues regarding sensitivity can be addressed[b]
PET scanning for staging locally advanced non small-cell lung cancer [310] (29)	OCOG	RCT	Establish clinical utility in making radical treatment decisions	Study terminated prematurely by efficacy and safety committee	Open-ended access to PET insured for this indication[a]
PET scanning for staging non small-cell lung cancer [322] (13)	OCOG	RCT	Resolve inconsistencies from 2 studies to inform decision to re-access	PET reduces futile thoracotomy rates	Open-ended access to PET insured for this indication[a]
PET scanning for staging breast cancer [320] (22)	OCOG	Prospective cohort	Compare PET to sentinel lymph node biopsy in staging breast cancer	No utility in staging breast cancer patients	Not insured[c]
PET scanning for colorectal cancer metastatic to liver	OCOG	RCT	Assess clinical utility of PET in decision for metastatectomy	Accrual completed February 2010	Awaiting results of results
PET scanning for head and neck cancer [400]	OCOG	Single arm prospective cohort	Assess PET as a decision tool prior to surgery following radiation therapy	Preliminary results fail to demonstrate clinical utility	Not insured if preliminary results are validated[c]
Extracorporeal photopheresis (EP) [120]	Single AHSC	Prospective observational	Effectiveness of EP in graft v host (GvH) disease and T-cell lymphoma to establish whether to develop a program	Effective for GvH disease, inconclusive for cutaneous T-cell lymphoma	Open-ended access approved for GvH disease[a] but continue to evaluate use in T-cell lymphoma[b]

[a]Widespread adoption.
[b]Limited adoption.
[c]Technology withdrawn.

reduced futile thoracotomy rates. While reduced thoracotomy rates attributed to PET were potentially cost-saving and posed reduced risks to some patients, missing potentially curative surgery due to low sensitivity or increasing futile thoracotomies due to low specificity, would be to the patient's detriment. A confirmatory RCT was deemed necessary and was conducted by OCOG (13).

Results. A total of 170 patients were assigned to PET-CT and 167 to conventional staging. Disease was correctly upstaged in 23 of 167 PET-CT recipients and 11 of 162 conventional staging recipients (13.8 percent versus 6.8 percent; [95 percent CI, 0.3 to 13.7 percentage points]). Disease was incorrectly upstaged in eight PET-CT recipients and one conventional staging recipient (4.8 percent versus 0.6 percent; difference [CI, 0.5 to 8.6 percentage points]), and it was incorrectly under-staged in 4.9 percent versus 29.6 percent [CI, 5.7 to 23.4 percentage points]). At 3 years, fifty-two patients who had PET-CT and fifty-seven patients who had conventional staging had died (13).

Preoperative staging with PET-CT identified more patients with mediastinal and extrathoracic disease than conventional staging, thereby sparing more patients from stage-inappropriate surgery. While the strategy also incorrectly upstaged disease in more patients, on balance, the benefits of performing PET scans outweighed the risks.

Policy Implications. Full funding was made available for PET scanning for this indication

PET in Locally Advanced NSCLC

Rationale. In locally advanced NSCL cancer, the clinical utility of PET to prevent aggressive treatment (combined modality [CMT] using radical radiation therapy (RT) plus chemotherapy) in patients upstaged by PET was unknown. Without an understanding of clinical utility, there was a concern that patients could have been inappropriately treated based on the results of a PET scan. Patients who were considered candidates for CMT were randomized to PET-CT or CT for RT treatment planning (29). The primary outcome was the proportion of patients who did not receive CMT because their tumor was upstaged or intrathoracic tumor was too extensive for radical RT.

Results. Following a planned interim analysis for the primary outcome, the Data Safety Monitoring Board recommended stopping recruitment because of superior efficacy with PET-CT. A total of 289/304 randomized patients had analyzable data. Twenty-five patients were unsuitable for CMT: twenty-one in the PET-CT arm (sixteen upstaged to Stage 4 and 5 unsuitable for radical RT) and four in the CT arm (unsuitable for radical RT) ($p = .0002$).

PET-CT was superior to CT alone in selecting appropriate patients for CMT. Longer follow-up will determine the impact on overall survival.

Policy Implications. Based on this CED, full funding was made available for PET scanning for this indication.

PET as an Adjunct in Pre-Operative Assessment for Liver Metastasectomy in Colon Cancer

Rationale. There was uncertainty regarding the clinical utility of PET as an adjunct to CT in the pre-operative assessment of patients undergoing liver metastasectomy. A prospective RCT of PET/CT versus CT alone was funded to resolve this issue before committing to a definitive funding decision.

Results. A total of 410 patients randomized and study completed. Results of the analysis are awaited.

Policy Implications. Definitive funding for this indication will be predicated on the results of this CED.

PET as an Adjunct to Conventional Imaging with CT to Aid Decision Whether to Proceed to Radical Neck Resection Following Radiation Therapy

Rationale. Surgery is undertaken for patients with locally advanced squamous cell carcinoma of the head and neck following radiation therapy with or without chemotherapy. There was uncertainty regarding the utility of PET scanning to improve patient selection for surgical neck dissection following RT. A cohort CED study was undertaken in which all patients underwent conventional imaging followed by PET scanning after receiving conventional RT. All patients subsequently underwent neck dissection and the results of PET scanning compared with pathology results.

Results. Following accrual of 400 patients, this study demonstrated that the addition of PET to conventional CT scanning before neck dissection had no clinical utility in determining whether patients should proceed to neck dissection. Detailed results await publication.

Policy Implications. Funding for this indication will be withdrawn according to expectations set out for the CED predicated on final results.

PET in Early Stage Breast Cancer

Rationale. Sentinel lymph node (SLN) biopsy has become an important investigation to aid decision making whether to undertake an axillary lymph node dissection for patients with early stage breast cancer. A cohort CED was funded to address the uncertainty as to whether PET scanning could replace or act as an adjunctive imaging technology for this purpose.

Results. A total of 336 women were entered into a cohort study and all received a PET scan before proceeding to sentinel lymph node examination. Sentinel nodes were found for 312/325 women (22). Using logistic regression, age, body mass index, number of nodes, and tumor size were assessed

as predictors of prevalence, positive PET, and sensitivity. Only tumor size was predictive ($p < .05$) for prevalence (OR = 1.6), PET positivity (OR = 1.7), and PET sensitivity (OR = 1.4). A low sensitivity and high positive predictive value for detection of axillary nodal metastases for PET compared with SLN biopsy suggested that this technology has limited, if any utility in staging breast cancer patients.

Policy Implications. Funding for this PET indication was withdrawn.

PET Registry

A PET registry was developed for indications where clinical trials were not feasible, where existing imaging modalities were unhelpful in the presence of collateral evidence of disease progression such as rising tumor markers with a normal or uncertain CT or MRI, or where good quality evidence was already available (e.g., solitary pulmonary nodule) but federal licensing precluded the used of the isotope fluoro-deoxy glucose (FDG) outside of a clinical trial.

Policy Implications. All PET studies for which evidence of clinical utility was demonstrated and registry studies were approved for funding as insured services in October 2009, comprising nine cancer and one cardiac indication. Additional approvals are anticipated in the future for indications in which clinical utility can be demonstrated.

Drug Eluting Stents and CT Angiography CED to Address Uncertainty Regarding Generalizability

The performance and use of a technology in the real world may differ from an RCT. Once diffused, adjunctive treatments excluded in the RCT may be provided with a new technology, or the technology may be used at a different stage of the disease than occurred in the RCT. The technology may not perform in the same way in a less-restrictive environment and especially when eligibility criteria are relaxed. The definition of successful outcomes may also not be as rigorous in the real world.

Drug Eluting Stents

Rationale. RCT evidence indicated that in patients with low risk coronary artery disease, the use of drug eluting stents (DES) resulted in significantly decreased restenosis rates when compared with patients who received conventional bare metal stents (BMS) (17;18). Based on these studies, acuity creep to high risk coronary artery disease seemed inevitable. Before making a definitive funding decision for DES, the MOHLTC funded PATH and ICES to undertake a prospective pragmatic study of patients who received BMS or DES to test the generalizability of this finding in the context of the Ontario health system. This was initiated soon after the RCT evidence was published.

Results. In this pragmatic prospective study, 23,000 patients were prospectively tracked on a modified existing patient outcomes and administrative database for BMS over an 18-month period to also cover DES (26). Propensity scores were used to find matched patients treated with DES or BMS. This was possible because of the large number of subjects, the collection of appropriate variables in the registry and the linkage to robust administrative data sets. Restenosis rates were similar for BMS and DES for low-risk patients, whereas high risk patients with narrow or long lesions with diabetes were found to derive benefit from DES (26).

Policy Implications. Based on the CED results, funding for DES was confined to high risk patients with diabetes, narrow and/or long stenotic coronary arteries. This led to a 36 percent conversion rate from BMS to DES in Ontario compared with an estimated 90 percent conversion rate in the United States (8) and an estimated annual saving of CAN$20 million. Possible reasons for difference in results included less restrictive criteria for use, the use of clopidogrel in stented patients, and the recent adoption of a newer version of the bare metal stent (personal communication).

CT Angiography

Rationale. CT angiography is a less invasive alternative to coronary angiography for the anatomical detection of coronary artery disease. A recent systematic review reported a pooled sensitivity of 0.96 (95 percent confidence interval [CI], 0.94–0.98), which was higher than for functional imaging (15). Uncertainty regarding clinical utility could result in over-use of this less-invasive anatomical diagnosis, duplication (many patients would still require a coronary angiogram), and creep to include asymptomatic individuals.

A CED was subsequently conducted by PATH in collaboration with leading cardiologists and radiologists to assess whether the published sensitivities were generalizable to Ontario to better inform decisions regarding the adoption of this technology. It was explicitly stated at the beginning of this study that a final decision regarding insurability would be predicated on the results of the CED. Patients referred for coronary angiography were invited to participate in this study in which they would undergo CT angiography before coronary angiography. Images from the two technologies were compared using dual reads and arbitration read to resolve disagreement.

Results. The sensitivity for CT angiography conducted at four academic health science centers was 82 percent, which was consistent with sensitivities reported for functional imaging modalities and lower than what was expected from published clinical trials (15). This study did not completely reflect real-world conditions as it was conducted with quality assurance and with experienced radiologists doing dual reads. It is expected that with more widespread adoption, sensitivity rates will be lower. Detailed results await publication (4).

Policy Implications. CT angiography remains uninsured and under policy consideration, which will be significantly affected by the results from this CED.

CED to Address Uncertainty Regarding Safety, and Indications – Endovascular Abdominal Aortic Aneurysm Repair

Rationale. A systematic review on the effectiveness and safety of abdominal aortic aneurysm repair (EVAAR) (14) concluded that there was evidence of effectiveness but uncertainty regarding endoleak from the edges of the deployed device. This was a potentially important technology as it offered a minimally invasive alternative to open abdominal surgery.

It was decided to undertake a CED to assess the safety of this device and its cost-effectiveness to better inform definitive funding decisions. A prospective observational single center study involving 160 patients and organized by PATH was undertaken (25).

Results. Type 2 endoleak was identified as a complication, but this type of endoleak was reported not to present a safety hazard or reduce long-term effectiveness. The economic analysis based on the study showed that substituting EVAAR for open surgery in high surgical-risk patients was cost-effective, but not so for patients at low risk for open surgical repair.

Policy Implications. Based on this study, increased funding for EVAAR was approved for patients at high risk for open surgical repair of the abdominal aortic aneurysm. Provincial funding for patients at low risk for open surgical repair has not been provided.

In another example of uncertainty regarding the use of a technology, a CED was undertaken to evaluate extracorporeal photopheresis in the treatment of cutaneous T-cell lymphoma and in graft versus host disease. Although the results of this study have not yet been published, they were used to make a decision to fund this intervention for graft versus host disease and continued study in patients with cutaneous T-cell lymphoma.

Uncertainty Regarding Long-Term Costs and Effects for Chronic Disease

Rationale. Planning for the uptake of health technologies for chronic diseases can be complex due to long time horizons, difficulties understanding events avoided by the introduction of these technologies and associated impacts on cost-effectiveness.

Using a diabetes outcomes model developed by Clarke et al. (5), PATH developed an Ontario-specific version of this micro simulation model. The model has been used repeatedly to inform policy. One example was its use as a core evaluation tool to assess components of the provincial diabetes strategy which provided a basis for prioritizing four major components of the strategy based on cost-effectiveness and avoidance of downstream complication rates over a 40-year time horizon. The use of this model formed part of a broad evidence-based approach in conjunction with a mega-analysis on these interventions (16). *Mega-analysis* is a term used at MAS in which technologies around a disease condition are disaggregated for evidence-based analysis and then re-aggregated around common outcomes to identify technologies most likely to optimize patient outcomes.

Results. Using estimates of common outcomes attributed to multidisciplinary care, insulin infusion pumps in the treatment of type 2 diabetes, behavior modification, and bariatric surgery in the management of morbidly obese individuals with type 2 diabetes, the most cost-effective strategy was bariatric surgery, followed by multidisciplinary care, and behavior modification, all of which had an incremental cost-effectiveness of under CAN$36,000. Insulin infusion pumps in insulin-dependent type 2 diabetes were not cost-effective. Whereas behavior modification interventions were cost-effective, multidisciplinary teams are trained in these techniques so this was not considered as a separate decision point.

Policy Implications. Following this analysis, CAN$110 million was allocated to improve bariatric surgery services and to encourage its use especially in patients with type 2 diabetes. Funding was also provided to increase the number of multidisciplinary teams. Insulin infusion pumps in patients with type 2 diabetes requiring insulin may not be pursued as a policy option.

THE EFFECTS OF FIELD EVALUATIONS ON SUBSEQUENT ADOPTION

The effects of CED results on the adoption of technology could be assessed for twelve interventions and one subgroup. The CED was responsible for widespread access to and adoption of technologies in six instances, limited access in three instances, and withdrawal of the technology in four instances. Although a formal quantitative analysis of changes in usage/diffusion was beyond the scope of this study, the general diffusion effects are summarized in Table 1 in which the effects of the field evaluation on diffusion of the technology are summarized in the last column as (a) for widespread adoption, (b) for limited adoption, and (c) for withdrawal of funding or removal of the technology.

DISCUSSION

Primary research to address uncertainty that persists after a formal review of existing evidence is receiving increasing attention (2;19;27;28) but experience is limited. The Ontario Field Evaluation Program and specifically the CED studies reported here demonstrate the breadth of studies that can be used to address uncertainty following systematic review and represents the largest body of evidence and experience to

https://doi.org/10.1017/S0266462311000018 Published online by Cambridge University Press

date. The ability of HTA to directly influence policy decision making is an important indication of its relevance. By this measure, we would conclude that post-market evaluation to address uncertainty following systematic review adds a meaningful dimension to HTA. Furthermore, CED makes an important contribution to definitive decision making based on the true performance of technologies under "real-world" conditions.

We have demonstrated that the performance of health technologies in the "real-world" setting may differ from outcomes derived from the more rigid constructs that RCTs must use to eliminate bias. While this should not detract from the essential role of RCTs in demonstrating efficacy, post-market studies make an important contribution to the body of evidence and may be necessary before making long-term funding commitments. This is especially when there is residual uncertainty regarding effectiveness, cost-effectiveness, or safety following systematic review. Ideally, post-market studies should not restrict access to the technology while they are being conducted, but there must be an understanding that definitive funding will be predicated on the results of these studies.

Traditionally cost-effectiveness has not been a major decision point for approving non-drug health technologies in Ontario. OHTAC initially used clinical evidence as the primary decision-making criterion. As a result, residual uncertainty that led to field evaluations was almost always within the domains of effectiveness or safety. This has changed over time and cost-effectiveness is now one of the four considerations in the OHTAC decision determinants model used for making its recommendations (11). After the adoption of this broader decision determinants framework, residual uncertainty could reside in one of several areas—effectiveness, efficiency, social/ethical values, or system feasibility. Over time, economic considerations in particular have assumed greater weight in the decision process, such as occurred in the CED studies on EVAAR for low operative risk patients, in four components of the diabetes strategy and in CT angiography.

The question remains why health systems should become involved in improving the quality of evidence of effectiveness and cost-effectiveness. Ignoring this opportunity at the inflection point of the diffusion curve may invite passive diffusion and intuitive decision making. Results from the completed CEDs raises questions regarding the future scope for CED. Our results demonstrate how the performance of health technologies under "real-world" conditions (effectiveness) can differ from reported RCT results (efficacy). In some situations, the lack of external validity from efficacy RCTs may mean that definitive funding decisions should not necessarily be based on this evidence alone and that CED should be considered for technologies to test generalizability from RCT data before making definitive funding decisions.

We emphasize the importance for collaborations among decision makers, academia, physicians, and institutions to conduct field evaluations. There is a tension between conducting these evaluations expeditiously and ensuring the evaluations are methodologically rigorous and defensible. We have found that evaluations can be facilitated through the use of health administrative databases and through the use of policy models which can be applied to various technology considerations.

This work is also facilitated by the ability to link comprehensive health administrative data sets to strengthen evidence-based analyses, field evaluations, and economic modeling. Furthermore, if the scope of CEDs expands, greater efficiencies will need to be introduced through inter-jurisdictional collaboration, collaboration with industry partners, and by undertaking some studies in the pre-marketing phase.

Not reported in this report are non-CED post-market evaluations undertaken by the Ontario Field Evaluation Program. These have included polling people to establish prevalence and acuity for stress urinary incontinence to plan for the adoption of mid-urethral slings and to study barriers to access for patients with diabetes; safety studies by the UHN Usability and Human Factors Laboratory of newer versions of infusion pumps, of portable in-room air cleaners and of CT scanners and MRI; and the use of human papilloma virus detection as an adjunct to cytological screening for cervical cancer.

Decision making under uncertainty is a complex process that goes beyond evidence of effectiveness and cost-effectiveness. Fiscally constrained health systems will increasingly look toward post-marketing studies to support funding decisions as they raise their threshold for tolerating uncertainty and seek ways to prioritize funding for competing needs. Probabilistic findings of medical effectiveness are fundamentally ambiguous as they relate to action (24). This means that policy makers will need a range of different types of answers when confronting a set of findings from an evidence-based review and may require different types of supplementary or field analyses based on their context.

It is important to decide how to prioritize which technologies should be subject to post-market evaluation. Value of information analysis can be used not only to set research priorities (6), but can and is being used in Ontario to define data collection for specific field evaluation studies (12).

As the scope of field evaluations including CED increases, we note the following methodological issues:

CED studies must be methodologically rigorous and defensible to be accepted by the research and health professional community. However, these studies must also be relevant to decision makers and this requires timely completion. Methods to increase patient accrual to RCTs and other prospective observational studies must be explored and the use of patient outcomes-linked administrative databases should be considered when possible.

Post-market evaluations should not unreasonably impede access to the technology being assessed.

Policy makers and opinion leaders should be involved in the formative process of field evaluations to ensure relevance of the outcome to decision making and to influence adoption of the technology according to the results.

CONCLUSION

The evaluation of health technologies under real-world conditions before adoption makes an important contribution to understanding their relevance to sustainability in health care. We have demonstrated that investment of this activity by health systems provides a sound basis for long-term funding decisions and can affect the adoption and associated costs for health technologies. Based on our experience, CED evaluation before making long-term funding decisions for widespread adoption, especially where there is uncertainty of effectiveness, safety, or cost-effectiveness, should be increasingly funded by health systems.

CONTACT INFORMATION

Leslie Levin, MD (Les.Levin@ontario.ca), Professor, Department of Medicine, University of Toronto, 21 King's College Circle, Toronto, Ontario M5T 1P7, Canada; Head, Medical Advisory Secretariat, Ministry of Health and Long-Term Care, 10th floor, 20 Dundas Street West, Toronto, Ontario M5G 2C2, Canada

Ron Goeree, MA (goereer@mcmaster.ca), Associate Professor, Department of Clinical Epidemiology and Biostatistics, McMaster University; Director, Programs for Assessment of Technologies in Health (PATH) Research Institute, St. Joseph's Healthcare Hamilton, 25 Main Street West, Suite 2000, Hamilton, Ontario L8P 1H1, Canada

Mark Levine, MD, MSc (mlevine@mcmaster.ca), Professor and Chair, Department of Oncology, McMaster University, 1200 Main Street West, Hamilton, Ontario L8N 3G5, Canada; Director, Ontario Clinical Oncology Group, c/o Henderson Research Centre, 711 Concession Street, Hamilton, Ontario L8V 1C3, Canada

Murray Krahn, MD, MSc (murray.krahn@theta.utoronto.ca), Professor and F. Norman Hughes Chair, Department of Medicine, University of Toronto, 21 King's College Circle, Toronto, Ontario M5T 1P7, Canada; Director, Faculty of Pharmacy, Toronto Health and Technology Assessment Collaborative, 144 College Street, Room 600, Toronto, Ontario M5S 3M2, Canada

Tony Easty, PhD, PEng, (tony.easty@uhn.on.ca), Associate Professor, Institute of Biomaterials and Biomedical Engineering, University of Toronto, 21 King's College Circle, Toronto, Ontario M5T 1P7, Canada; Senior Scientist, Centre for Global eHealth Innovation, University Health Network, 190 Elizabeth Street, Toronto, Ontario M5G 2C4, Canada

Adalstein Brown, PhD (adalsteinn.brown@utoronto.ca), Dalla Lana Chair in Public Health Policy, Department of Public Health Policy, Dalla Lana School of Public Health, University of Toronto, Sixth Floor, 155 College Street, Toronto, Ontario M5T 3M7, Canada; Scientist, Li Ka Shing Knowledge Institute, St. Michael's Hospital, 30 Bond Street, Toronto, Ontario M5B 1W8, Canada

David Henry, MB, CB (David.Henry@ICES.on.ca) Professor, Department of Medicine, University of Toronto, Ontario, Canada; President and CEO, Institute for Clinical Evaluative Sciences (ICES), G106, 2075 Bayview Avenue, Toronto, Ontario M4N 3M5, Canada

CONFLICT OF INTEREST

L Levin has received travel funding from HTAi, Institute of Medicine, and the Korean HTA agency NSCR. M Krahn has received a grant from Johnson & Johnson. T Easty's institute has received a grant from the Ministry of Health in Ontario for this work and other funding from NSERC, CIHR, CPSI, and Hospira Corp. The other authors do not report any potential conflicts of interest.

REFERENCES

1. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ.* 2004;328:1490.
2. Chalkidou K, Hoy A, Littlejohns P. Making a decision to wait for more evidence: When the National Institute for Health and Clinical Excellence recommends a technology only in the context of research. *J R Soc Med.* 2007;100:453-460.
3. Chalkidou K, Tunis S, Lopert R, et al. Comparative effectiveness research and evidence-based health policy: Experience from four countries. *Milbank Q.* 2009;87:339-367.
4. Chow BJW, Freeman MR, Bowen JM, et al. For the OMCAS Investigators. Ontario Multi-Director Computed Tomographic Coronary Angiography Study (OMCAS): A field evaluation of diagnostic accuracy. *Arch Intern Med*. (in press).
5. Clarke PM, Gray AM, Briggs A, et al. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: The United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). *Diabetologia.* 2004;47:1747-1759.
6. Claxton K, Briggs A, Buxton MJ, et al. Value based pricing for NHS drugs: An opportunity not to be missed? *BMJ.* 2008;336:251-254.
7. Evans WK, Laupacis A, Gulenchyn KY, Levin L, Levine M. Evidence-based approach to the introduction of positron emission tomography in Ontario, Canada. *J Clin Oncol.* 2009;27:5607-5613.
8. Feder BJ. *Doctors rethink widespread use of heart stents*. October 21, 2006. The New York Times. http://www.nytimes.com/2006/10/21/business/21stent.html (accessed December 2, 2009).
9. Goeree R, Levin L. Building bridges between academic research and policy formulation: The PRUFE framework – an integral part of Ontario's evidence-based HTPA process. *Pharmacoeconomics.* 2006;24:1143-1156.
10. Institute for Clinical Evaluative Sciences. *Health technology assessment of positron emission tomography (PET) – A systematic review*. May 2001. http://www.ices.on.ca/file/Health_

Technology_Assessment-PET_May-2001.pdf (accessed December 2, 2009).
11. Johnson AP, Sikich NJ, Evans G, et al. Health technology assessment: A comprehensive framework for evidence-based recommendations in Ontario. *Int J Technol Assess Health Care.* 2009;25:141-150.
12. Levin L, Goeree R, Sikich N, et al. Establishing a comprehensive continuum from an evidentiary base to policy development for health technologies: The Ontario experience. *Int J Technol Assess Health Care.* 2007;23:299-309.
13. Maziak DE, Darling GE, Inculet RI, et al. Positron emission tomography in staging early lung cancer: A randomized trial. *Ann Intern Med.* 2009;151:221-248.
14. Medical Advisory Secretariat. *Endovascular repair of abdominal aortic aneurysm: An evidence-based analysis*. Ontario Health Technol Assess Ser. [Internet]. 2002;2:1-46. http://www.health.gov.on.ca/english/providers/program/mas/tech/reviews/pdf/rev_evar_030102.pdf.
15. Medical Advisory Secretariat. *Non-invasive cardiac imaging technologies for the diagnosis of coronary artery disease: A summary of evidence-based analyses*. Ont Health Technol Assess Ser. [Internet]. 2010;10:1-40. http://www.health.gov.on.ca/english/providers/program/mas/tech/reviews/pdf/cardiac_aggregate_20100527.pdf.
16. Medical Advisory Secretariat. Diabetes strategy evidence platform: A summary of evidence-based analyses. *Ont Health Technol Assess Ser.* 2009;9:1-43. http://www.health.gov.on.ca/english/providers/program/mas/tech/reviews/pdf/rev_diabetes_sum_20091020.pdf.
17. Morice MC, Serruys PW, Sousa JE, et al. A randomized comparison of a sirolimus-eluting stent with a standard stent for coronary revascularization. *N Engl J Med*. 2002;346:1773-1780.
18. Moses JW, Leon MB, Popma JJ, et al. Sirolimus-eluting stents versus standard stents in patients with stenosis in a native coronary artery. *N Engl J Med*. 2003;349:1315-1323.
19. National Institute for Health and Clinical Excellence. *Research and development strategy*. London: National Institute for Health and Clinical Excellence; 2004. http://www.nice.org.uk/niceMedia/pdf/R&D_strategy_2006.pdf. (accessed December 2, 2009).
20. Neumann PJ, Tunis SR. Medicare and medical technology—the growing demand for relevant outcomes. *N Engl J Med.* 2010;362:377-379.
21. O'Reilly D, Hopkins R, Blackhouse G, et al. Long-term cost-utility analysis of a multidisciplinary primary care diabetes management program in Ontario. *Can J Diabetes.* 2007;31:205-214.
22. Pritchard KI, Julian J, McCready D. A prospective study evaluating 18F-Fluorodeoxyglucose (18FDG) positron emission tomography (PET) in the assessment of axillary nodal spread in women undergoing sentinel lymph node biopsy (SLNB) for breast cancer. *J Clin Oncol.* 2008;26 (Suppl):533.
23. Stadtmauer EA, O'Neill A, Goldstein LJ, et al. Conventional-dose chemotherapy compared with high-dose chemotherapy plus autologous hematopoietic stem-cell transplantation for metastatic breast cancer. Philadelphia Bone Marrow Transplant Group. *N Engl J Med.* 2000;342:1069-1076.
24. Tanenbaum SJ. "Medical effectiveness" in Canadian and U.S. health policy: The comparative politics of inferential ambiguity. *Health Serv Res.* 1996;31:517-532.
25. Tarride JE, Blackhouse G, De Rose G, et al. Cost-effectiveness analysis of elective endovascular repair compared with open surgical repair of abdominal aortic aneurysms for patients at a high surgical risk: A 1-year patient-level analysis conducted in Ontario, Canada. *J Vasc Surg.* 2008;48:779-787.
26. Tu JV, Bowen J, Chiu M, et al. Effectiveness and safety of drug-eluting stents in Ontario. *N Engl J Med.* 2007;357:1393-1402.
27. Tunis SR, Pearson SD. Coverage options for promising technologies: Medicare's 'coverage with evidence development'. *Health Aff (Millwood).* 2006;25:1218-1230.
28. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: Increasing the value of clinical research for decision making in clinical and health policy. *JAMA.* 2003;290:1624-1632.
29. Ung Y, Sun A, MacRae R. Impact of positron emission tomography (PET) in stage III non-small cell lung cancer (NSCLC): A prospective randomized trial (PET START). *J Clin Oncol.* 2009;27: 15(Suppl):7548.
30. Whicher DM, Chalkidou K, Dhalla IA, Levin L, Tunis S. Comparative effectiveness research in Ontario, Canada: Producing relevant and timely information for health care decision makers. *Milbank Q.* 2009;87:585-606.