Canada's Drug Agency
L'Agence des médicaments du Canada

March 2025

Drugs **Health Technologies** Health Systems

Methods Guide for Health Technology Assessment

Foreword

This methods guide describes the methods involved in conducting health technology assessment (HTA) at Canada's Drug Agency (CDA-AMC). This methods guide will focus on the appraisal of the clinical evidence for a drug product submitted by a sponsor to CDA-AMC to address the core HTA research question of comparative effectiveness and harms. The goals in development of this guide were as follows:

- to highlight the types of clinical evidence that can inform the comparative effectiveness and harms of a drug product
- to identify key methods and their use in the evaluation of clinical evidence for drug products submitted to CDA-AMC for HTA
- to facilitate the generation and reporting of the clinical evidence by the sponsors
- to be transparent in how CDA-AMC reviewers appraise and report on the assessment of the clinical evidence.

This methods guide was developed by first identifying core topics, collating and reviewing methods and best practices internationally, and then selecting the appropriate methods as applicable to HTA conducted in Canada for inclusion. This guide leverages existing methods documents from other provincial, national, and international regulatory and HTA agencies, which are cited throughout. We consulted with and received feedback from technical experts and other relevant parties, including Health Canada, l'Institut national d'excellence en santé et en services sociaux (INESSS), international regulatory and HTA agencies, representatives from the pharmaceutical industry and others who hold or generate data, clinicians, patient organizations, and the general public. Modifications were then made to the draft based on the feedback received.

This guidance is intended for use by those who generate and submit evidence, and those who conduct the evidence appraisal. The aim is to be iterative and periodically update or add to the guidance over time as methods evolve or to address emergent issues.

Table of Contents

Abbreviations

AE	adverse event
CDA-AMC	Canada's Drug Agency
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HTA	health technology assessment
IPD	individual patient data
ITC	indirect treatment comparison
LTE	long-term extension
MAIC	matching-adjusted indirect comparison
MID	minimal important difference
NMA	network meta-analysis
OS	overall survival
PFS	progression-free survival
PICO(T)(S)	population(s), intervention(s), comparator(s), outcome(s), time or time frame, study design or setting
RCT	randomized controlled trial
ROBINS-I	Risk of Bias in Nonrandomized Studies of Interventions
RWD	real-world data
RWE	real-world evidence
SR	systematic review

Authors and Contributors

The methods guide was developed and drafted by an internal working group led by members of the Science and Methods, Quantitative Methods team:

- Matthew Bryan, PhD
- Saunya Dover, MSc
- Allison Gates, PhD
- Michelle Gates, PhD
- Mary-Ellen Hogan, PharmD, PhD
- Farah Husein, PharmD, MSc
- Mike Innes, MSc
- Eric Morenz, PhD, MSc
- Hongbo Yuan, PhD

Contributions were made by representatives from cross-functional teams at CDA-AMC:

- Ethics: Renata Axler, PhD
- Medical Devices and Clinical Interventions: Angela Barbara, PhD, MSc
- HTA and Deliberation: Deirdre DeJean, PhD
- Qualitative Methods: Viviane Grandpierre, PhD
- Health Economics: Alex Haines, MSc
- Pharmaceutical Reviews: Gaetanne Murphy, MSc
- Scientific Advice: Sheri Pohar, PhD
- Scientific Advice and Real-World Evidence: Colette Raymond, PharmD
- Scientific Advice: Amy Sood, PhD
- Post-Market Drug Evaluation: David Stock, PhD

Introduction

Health Technology Assessment at Canada's Drug Agency

The joint task force of the International Network of Agencies for Health Technology Assessment and Health Technology Assessment International defines health technology assessment (HTA) as "a multidisciplinary process that uses explicit methods to determine the value of a health technology at different points in its life cycle. The purpose is to inform decision-making in order to promote an equitable, efficient, and high-quality health system."[1]

HTA at Canada's Drug Agency (CDA-AMC) involves the development of reports that present appraisals of the evidence: **clinical evidence** on the benefits and harms of a drug or medical device relative to relevant comparators used in clinical practice across Canada; **economic evidence** on cost-effectiveness and long-term value relative to other standard of care options; and **value considerations** including — but not limited to — unmet needs, patients' experiences and values, clinicians' experiences and perspectives, perspectives of health care systems (including clinical experts and public drug plans), social and ethical considerations, health equity, and other relevant considerations. The HTA reports are 1 of several inputs to the expert committees that inform nonbinding recommendations, which are subsequently used in reimbursement decision-making by Canada's federal, provincial, and territorial governments, with the exception of Quebec.

This methods guide focuses on the methods related to the **clinical evidence** for a drug product submitted by a sponsor to CDA-AMC to address the core HTA questions of comparative effectiveness and harms, and long-term value. While this methodological approach will apply to most drug product submissions and resubmissions to CDA-AMC — referred to here as the reference case — unique circumstances for specific drugs, indications, or other factors may warrant alternative approaches, which would be determined on a case-by-case basis. This methods guide does not apply to clinical evidence reviews in which the primary question does not pertain to comparative effectiveness and harms (or to interventions that are not standalone drug products, such as drug-device combinations or companion diagnostics).

References are made to various tools, checklists, guidance documents, and other sources throughout this methods guide. For each review, CDA-AMC considers which method(s) and tool(s) are most appropriate to assess the clinical evidence, with documentation of any choice(s) by the reviewer in the clinical report.

It should be noted that while the appraisal of the clinical evidence informs the deliberation of the respective expert committee, this is only one of the considerations (along with economic evidence and value considerations) in the decision-making process and committee recommendation.

Assessment of the Clinical Evidence

Research Question and Scope

There are 3 core concepts regarding testing health care interventions in evidence-based medicine.

- Efficacy: Can it work?

- Effectiveness: Does it work in practice?
- Value: Is it worth it?[2]

The scope of the CDA-AMC Clinical Review report is informed by the HTA question of interest. For the reference case — the assessment of the clinical evidence for a drug product submitted to CDA-AMC by a sponsor — the main HTA question of interest is as follows:

> What are the effectiveness and harms of the drug product under review relative to relevant comparators used in clinical practice in Canada?

Efficacy describes how an intervention produces the desired effects under ideal, controlled conditions typically assessed in clinical trials designed to prioritize validity of inference within the trial population. Effectiveness describes how an intervention performs in real-world settings with study designs and estimands that prioritize an alignment between the sampled population and the target population.[62] Harms refers to both adverse events (AEs) and adverse drug reactions (side effects) of the drug product under review.

Relevant comparators are treatment alternatives used in the target population and clinical setting under review. Typically, these include other drugs used to treat patients in Canada for the same indication. Active comparators (i.e., nonplacebo interventions) are most relevant in the assessment of comparative effectiveness. In cases where there are no active treatment comparators, standard of care or best supportive care may also be considered as relevant comparators.

Another key consideration of the review is the evaluation of long-term comparative effectiveness and harms. Long-term extension (LTE) clinical studies are often open-label and single arm, thus not designed to generate direct evidence on comparative long-term-effectiveness and harms.

Sponsor submissions to CDA-AMC in support of a drug product should explicitly state how each component of the submitted clinical evidence informs the comparative effectiveness and potential harms of the drug product under review. For a drug product resubmission, the principles outlined in this methods guide will still apply, with the added requirement that the resubmission provides new evidence that also addresses any gaps or uncertainties identified during the initial review (unless the resubmission is for a subpopulation for which the evidence had previously been submitted to CDA-AMC).

The eligibility of studies for inclusion in the sponsor's clinical evidence submission should be informed by the scope of the review, the definition of relevant estimands, and defined by PICO(T)(S):

- population(s)
- intervention(s)
- comparator(s)
- outcome(s)
- time or time frame (if relevant)

- study design or setting (if relevant).

For the reference case, expectations for the sponsor's evidence submission are as follows.

Population

The population should be defined by the approved or proposed Health Canada indication. In some cases, CDA-AMC may consider a request from the sponsor to deviate from the Health Canada indication to align the population with that defined in the sponsor's reimbursement request and/or pharmacoeconomic analysis. Relevant population subgroups (e.g., those defined by age, sex, disease severity, or other characteristics) should be identified if they are pertinent to the reimbursement recommendation, or of interest to patients, clinicians, and/or payers (e.g., public drug programs). Subgroups included in the sponsor's pharmacoeconomic submission should also be specified as part of the sponsor's clinical evidence submission.

Interventions

The intervention under review by CDA-AMC is the drug product, including its formulation, dosage range(s), and route(s) of administration as approved or being considered by Health Canada.

Comparators

CDA-AMC considers clinically relevant comparators to include drug products used in clinical practice in Canada to treat patients described in the indication under review. These may include, but are not limited to:

- drug products with a Notice of Compliance for the indication under review and available (i.e., marketed) in Canada, and within the Health Canada–approved dosage range
- drug products that are not approved by Health Canada for the indication under review, if they are standard of care used in clinical practice in Canada and their use is supported by evidence, preferably evidence-informed clinical practice guidelines (including drug products available through Health Canada's Special Access Program)
- in rare circumstances, nondrug comparators that are used in clinical practice in Canada as interventions for the indication under review.

Sponsors should justify their selection of the relevant comparator(s) in their evidence submission package, including why selected drug product comparators are included and why other drug products used for that indication to treat patients in Canada are deemed not relevant. While the clinical evidence submission may include comparators not considered in the Pharmacoeconomic Review (e.g., comparators that are not publicly funded), all comparators used in the Pharmacoeconomic Review should be included in the clinical evidence submission.

Outcomes

Target outcomes of interest in HTA are those that estimate the clinical benefit and thereby help estimate the clinical value of the drug product. The assessment of health benefits considers clinically meaningful end points such as mortality; morbidity; and patient-reported experiences, symptoms, health behaviours, functioning, and health-related quality of life.[3] Target outcomes that are important to patients, clinicians, and/

or health system decision-makers include both effectiveness and harms outcomes (i.e., AEs deemed related to the drug product, serious adverse events [SAEs], withdrawals due to AEs, and death).

These "patient-relevant" final outcomes of interest for the CDA-AMC appraisal are selected from among the end points in the clinical evidence submitted by the sponsor. These are defined by CDA-AMC reviewers and clinical experts, by considering input from patients and patient groups, clinicians, and the public drug plans, as well as those identified in previous CDA-AMC or other HTA reviews relevant to the indication. CDA-AMC reviewers may also consider core outcome sets, using sources such as the Core Outcome Measures in Effectiveness Trials (COMET) database.[4]

Surrogate Outcomes

Surrogate outcomes are commonly defined[5] as biomarkers or intermediate outcomes that serve to substitute for patient-relevant final outcomes, and reliably predict benefits or harms based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence. This definition encompasses both biomarkers (such as blood pressure and tumour response) and intermediate outcomes (such as progression-free survival [PFS]) that may have potential direct relevance to patients.[6] A surrogate outcome is used in clinical trials when the direct measurement of patients' symptoms, functioning, quality of life or survival is impractical or infeasible.[6-8] A consequence of using surrogate end points is an increase in the uncertainty of an intervention's "true value" (including clinical efficacy or effectiveness, harms, and cost-effectiveness).[5] This uncertainty is reduced when surrogate end points are rigorously validated by statistical evidence that the treatment effect on a surrogate end point is strongly predictive of the treatment effect on the relevant target outcome.[3] The CDA-AMC reviewers may consult appropriate sources to support the use of specific surrogate outcomes, such as the Surrogate Outcome Table posted by the FDA.[9] The uncertainty around long-term value may be further mitigated by relevant real-world evidence (RWE) on validation, predictiveness, and correlation with patient-relevant outcomes.

For any use of a surrogate outcome, the clinical evidence underlying or justifying the use of a surrogate — as well as the evidence for the validation of the surrogate — should be explicitly stated in the sponsor's submission. If the surrogate outcome of interest is used in an indirect comparison, details on the surrogate outcome definition for each individual trial should also be included in the sponsor's submission. This evidence will be appraised by the CDA-AMC reviewers, in consultation with clinical experts and relevant literature as required.

Time

The duration of follow-up should be adequate to capture the outcomes defined for the trial. The timing of when the intervention is administered in the care or treatment pathway should also be considered and justified.

Study Design

In general, systematic reviews (SRs) on efficacy and/or effectiveness are generally limited to randomized trial designs. However, all trials submitted to Health Canada as pivotal, regardless of study design, as well as

nonpivotal phase III and IV randomized controlled trials (RCTs), should be included in the sponsor-submitted SR for appraisal by CDA-AMC.

The study designs best suited to inform comparative effectiveness and harms are outlined in the CDA-AMC reference case. When data from RCTs are not available or there are gaps in the clinical trial evidence, other studies may be included in the sponsor submission on a case-by-case basis, such as single-arm trials, LTE studies, indirect treatment comparisons (ITCs), and RWE studies. Additional details on relevant study designs and considerations for their appraisal can be found in subsequent sections.

Target Estimands

A well-defined estimand framework, incorporated into the methods reporting, enhances clarity and transparency in evidence evaluation. Study objectives are translated into the primary research question by defining an estimand, which is a detailed description of the treatment effect that a study sets out to quantify for a specified outcome.[10] Including estimand descriptions in study technical reports enables CDA-AMC to clearly understand the objectives of each individual study submitted for consideration, and evaluate how the study design, analysis, and estimation align with the specified estimand. Furthermore, this framework allows the CDA-AMC reviewers to determine whether the targeted estimand of each study addresses the Reimbursement Review question of interest from an HTA perspective.

Guidelines from the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) have specified 5 attributes for describing an estimand for study: the treatment or intervention of interest (and as appropriate, the alternative treatment or intervention to which comparison will be made), the population or patients targeted by the scientific question, the outcome of interest, specifications for handling of intercurrent events, and a population-level summary measure that is the basis for comparison.[10] However, when evaluating evidence for a drug product submitted for Reimbursement Review at CDA-AMC, additional attributes are considered important because such reviews often consider evidence that was not specifically developed for the purpose of informing reimbursement from an HTA perspective.[11] Evidence originally designed to inform regulatory decision-making often focuses on estimands that quantify the efficacy of an intervention relative to placebo or standard of care. In contrast, the HTA question of interest will generally require an estimand that evaluates the comparative effectiveness of an intervention relative to relevant active comparators. Thus, to promote consistency in HTA reporting, a standard framework has been proposed for reporting the target estimand of a study to an HTA entity, referred to as the population(s), intervention(s), comparator(s), outcome(s), summary effect measure, and intercurrent events (PICOSI) framework.[11] Here, intercurrent events refer to those events of interest that occur after treatment has been initiated that may impact the interpretation of the end point quantifying the treatment effect (e.g., the event modifies the treatment effect, such as in the use of rescue medications or other concomitant treatments, or has implications for adherence to the treatment regimen, including premature treatment discontinuation). CDA-AMC recommends reporting the target estimand for all comparative evidence submitted for reimbursement considerations — including RCTs, nonrandomized studies, and RWE — using this framework.[11]

In the context of RCTs, for each individual trial included in the clinical evidence submission to CDA-AMC, the target estimand for the primary end point(s) of the trial must be clearly identified in the technical report (e.g., study protocol, statistical analysis plan, or Clinical Study Report). The estimand(s) for the primary end point should be defined for the individual trial a priori (i.e., before data collection, analysis, or reporting). As outlined previously, key information should also include sufficient details of the intervention and the comparator(s) being evaluated to understand relevant intercurrent events and how these may impact the primary end point(s). For study designs other than RCTs, these same principles should be applied where possible. Sponsors may also choose to present results from a trial that targets alternative estimands, including those for secondary and exploratory end points and/or alternative analyses for the primary end point.

CDA-AMC defines the estimand of interest for informing reimbursement decisions, as the treatment effect describing the comparative effectiveness and harms of a drug product compared to all relevant comparators in clinical settings in Canada. This definition applies to any outcome(s) relevant to decision-making and assumes that all interventions are implemented as they are used or anticipated to be used in clinical practice in Canada. This will generally align with the disease indication as described in the product monograph submitted to and approved by Health Canada, and/or be consistent with established clinical practice guidelines or clinical practice in Canada.

Evidence Base

The clinical evidence submitted by the sponsor should address the HTA question and scope, as previously defined. The primary objective of the CDA-AMC review is to assess the drug product's comparative effectiveness and harms relative to relevant comparators in the treatment landscape in Canada. Sponsor submissions should include studies designed, conducted, and analyzed with methodological rigour to minimize any bias and confounding. The submission should also include transparent reporting of study design, analysis, data, and results.

Comparative effectiveness is informed by decision-grade evidence arising from multiple sources. Explanatory trials — and in particular multi-arm RCTs and rigorously conducted SRs of RCTs — are most likely to provide evidence that supports valid and causative scientific conclusions for the drug product under review. For this reason, the certainty of conclusions about the drug product under review from other study designs, including LTE studies and RWE, are generally lower than those able to be drawn from high-quality RCTs (refer to Table 1).

Table 1: Evidence Types for HTA Submissions

Evidence type	HTA purpose	Key strengths	Key limitations
RCT: multi-arm, with multiple relevant active comparators[a]	• Provides direct evidence of comparative clinical efficacy and harms of the drug product under review compared to	• Designed to generate robust evidence to potential biases, minimizes confounding, and isolates effects solely due to treatment, providing reliable and replicable estimates of	• Often not feasible to design, conduct, and/or report data in a timely manner • Sample population may not be representative of target population

Evidence type	HTA purpose	Key strengths	Key limitations
	multiple relevant active comparators	treatment effects • Facilitates valid causal inference for each relevant comparator within the sampled clinical setting	• Treatment landscape often evolves during the time to conduct the RCT; consequently, the RCT may not include comparators relevant to clinical practice in Canada • Long-term outcomes are often unavailable
RCT: at least 1 relevant active comparator[a]	• Provides direct evidence of comparative clinical efficacy and harms of the drug product under review relative to at least 1 relevant comparator	• Common evidence base type: designed to generate robust evidence to potential biases, minimizes confounding, and isolates effects solely due to treatment, providing reliable estimates of treatment effects within the sampled population • In some clinical settings, conducting multiple separate RCTs for each comparator may be more appropriate than a multi-arm RCT design	• Insights into comparative effectiveness and harms relative to other relevant active comparators may require indirect comparison methodologies • Sample population may not be representative of target population • Treatment landscape often evolves during the time taken to conduct the RCT; consequently, the RCT may not include comparators relevant to clinical practice in Canada • Long-term outcomes are often unavailable
RCT: placebo-controlled[a]	• Provides direct evidence of clinical efficacy and harms	• Common evidence base type: designed to generate robust evidence to potential biases, minimizes confounding, and isolates effects solely due to treatment, providing reliable estimates of treatment effects within the sampled population	• Insights into comparative effectiveness and harms relative to other relevant comparators requires indirect comparison methodologies • Sample population may not be representative of target population • Long-term outcomes are often unavailable
ITCs applied to RCTs and SRs of RCTs	• Synthesizes evidence using explicit and reproducible methods to systematically search for, select, critically appraise, and synthesize results of multiple primary studies	• Robust method to collate evidence from multiple primary studies to provide an objective summary of the balance of benefits and harms of an intervention for use by decision-makers • Can help to identify limitations of previous primary studies	• The ability to draw high-certainty conclusions can be equally reliant on the rigour of the ITC and SR methods and the conduct and characteristics of the included studies and the alignment between trial populations and target population
Single-arm, open-label (interventional) trials	• May provide some evidence of clinical efficacy and harms	• Commonly conducted in settings where an RCT is not feasible due to ethical or practical reasons (e.g., in rare	• Insights into comparative clinical efficacy and/or effectiveness and harms relative to relevant comparators

Evidence type	HTA purpose	Key strengths	Key limitations
		disease settings in which the number of available patients is less than what is required to power an RCT)	are typically of very low certainty, given methodologic limitations of external controls and indirect comparisons
RWE and other observational studies	• May be supplemental to clinical trial evidence or may be the primary evidence base for a subpopulation, new indication, or label expansion • May provide direct evidence of comparative effectiveness and harms relative to the relevant comparator(s), or may substitute an RCT when it is not feasible (e.g., orphan drug development with a single-arm trial combined with an external control arm) • May address evidence gaps on long-term effectiveness and harms	• Compared to RCTs, these study designs may enhance validity by aligning the sample population with the target population • Estimates of effectiveness can be measured within various patient populations, and under differing treatment strategies	• Vulnerable to confounding and other systematic sample biases • Typically, less able to produce robust estimates of treatment effects, especially in the presence of measurement error or missing data
Long-term extension (clinical) study	• Provides evidence on longer-term efficacy and harms	• Allows for the identification of side effects that may not have been observed during short-term use of interventions • Allows for the evaluation of durability of the treatment response	• Open-label and single-arm; often not designed to generate direct evidence on comparative long-term effectiveness and harms due to feasibility and other biases • Higher risk of missing data due to increasing attrition rates over time

HTA = health technology assessment; ITC = indirect treatment comparison; RCT = randomized controlled trial; RWE = real-world evidence; SR = systematic review.
[a]RCTs can have varying degrees of pragmatism; those that are more explanatory will have more limited generalizability, while those that are more pragmatic will have greater generalizability.

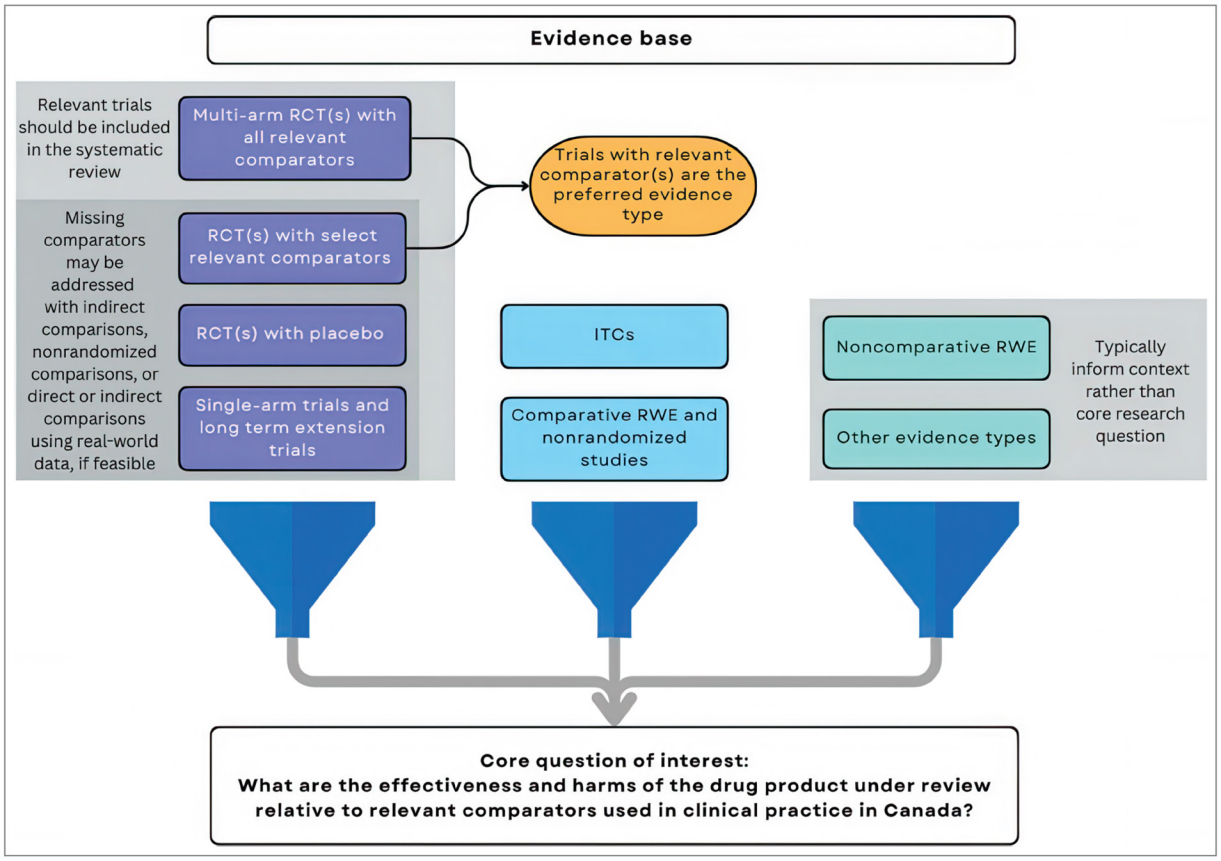
In the absence of direct head-to-head clinical trial evidence, indirect comparisons are acceptable alternatives for providing comparative evidence, provided that the methodology is appropriate, the indirect comparison is of sufficient quality, and methods and results are reported transparently. Indirect comparisons typically include, but are not limited to, network meta-analyses (NMAs) or population-adjusted indirect comparison methodologies such as matching-adjusted indirect comparisons (MAICs) (refer to Table 2).

Observational and RWE studies, including single-arm clinical trials in which there is an external control arm (historical control or contemporaneous cohort), may be considered on a case-by-case basis when it is not feasible or ethical to conduct an RCT, or when there are uncertainties in the clinical interventional trial and/or indirect comparison evidence. Long-term evidence on effectiveness and harms may come from LTE clinical

trials, or from observational evidence drawing upon real-world data (RWD) sources such as disease and drug registry data, administrative health data, and pharmacovigilance data.

The purpose, key strengths, and limitations of different study designs for the purpose of HTA evidence appraisals are outlined in Table 1. The general approach to the assessment of clinical evidence at CDA-AMC is outlined in Figure 1.

Figure 1: Appropriateness of Clinical Evidence Types for Inclusion in the CDA-AMC Appraisal to Address the HTA Question of Interest



ITC = indirect treatment comparison; RCT = randomized controlled trial; RWE = real-world evidence.

The framework depicted in Figure 1 is based on the relative strength of evidence of each evidence type as described in Table 1. The focus of the CDA-AMC assessment is based on the certainty of evidence for patient-relevant benefits and harms relative to relevant comparators, and the generalizability to clinical practice in Canada for the population specified by the indication of the drug product under review.

The following sections outline CDA-AMC methods for appraising the following evidence submitted by sponsors:

- **Comparative clinical efficacy or effectiveness and harms:** RCTs, SRs of RCTs, indirect comparisons

- **Long-term evidence:** LTE clinical studies

Refer also to the subsequent section on **RWE** and observational studies.

Evaluating the Evidence

Evidence Sources and Methods Overview

SR of Pivotal and Other Clinical Trials

Pivotal trials and RCTs are identified by the sponsor using a systematic search and selection procedure in adherence with relevant procedural guidance,[12] which includes transparency on the methods used. The review team at CDA-AMC completes subsequent steps of the SR, including data extraction and verification, critical appraisal, summary or synthesis (in the case of 2 or more submitted studies that are adequately similar in their PICOTS elements), and certainty of evidence assessment based on the studies submitted by the sponsor.

A single CDA-AMC reviewer abstracts the characteristics and results of the studies included in the sponsor-submitted SR, with independent verification by a second reviewer.

CDA-AMC recommends the inclusion of both absolute (e.g., mean difference, risk difference) and relative (e.g., risk ratio, odds ratio) between-group differences and confidence intervals to adequately interpret the clinical importance of effect estimates and to facilitate the certainty of evidence appraisal. For time-to-event outcomes, CDA-AMC requires estimates of between-group differences in event or event-free probabilities with 95% confidence intervals, which are often estimated using Kaplan-Meier curves, with accompanying data on the number of patients at risk at distinct time intervals, or hazard ratios from Cox regression analyses.[13] In the absence of this information, the CDA-AMC review team may not be able to fully assess the clinical importance of the estimated effect for a given outcome, and this uncertainty would be reflected in the clinical report.

When 2 or more submitted trials are adequately similar in PICO(T)(S), their outcome data are described together using a narrative summary,[14,15] where the size and direction of effect as well as the sample size and/or number of events of each contributing study are considered. The CDA-AMC review team does not synthesize study results statistically (e.g., via meta-analysis); however, statistical syntheses and integrated summaries of effectiveness submitted by the sponsor may be considered on a case-by-case basis and appraised according to relevant guidance.[16] Should discrepancies or inconsistencies be found in the data, the CDA-AMC review team will request clarification from the sponsor. The methods for synthesis should adhere to accepted methodological standards (e.g., FDA guidance for integrated summaries of effectiveness;[16] Cochrane guidance for meta-analysis)[17] and be informed by an a priori protocol. The methods should be reported in adequate detail to allow for critical appraisal by the CDA-AMC review team (e.g., in adherence to the Preferred Reporting Items for Systematic reviews and Meta-Analyses [PRISMA] 2020 reporting items).[18] In all cases, the characteristics and results of the individual studies contributing to

the synthesis should be reported alongside pooled estimates of effect. When multiple trials are submitted but are considered too dissimilar in their PICO(T)(S), their results are presented separately.

LTE Studies

Table 1 outlines the purpose, key strengths, and limitations of LTE studies.

LTE studies of pivotal trials and other RCTs are included in the CDA-AMC appraisal of the evidence when these have been completed, or interim analyses are available. Only LTE evidence submitted by the sponsor is included. When no evidence from LTE studies has been submitted by the sponsor, this is stated explicitly in the CDA-AMC clinical report.

Open-label extension studies are appraised using the same methodological approaches as outlined in the SR section: data abstraction, synthesis or summary of results, and interpretation of the clinical importance of effect estimates for relevant outcomes. The certainty of evidence is not formally assessed using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE)[19] approach (refer to the Certainty of Evidence section for more details), but relevant domains are considered when interpreting the findings.

Indirect Evidence

In the absence of head-to-head studies comparing 2 or more interventions of interest to provide direct evidence of comparative effectiveness and harms of the drug product under review, ITCs that follow internationally accepted reporting guidance and standards may be used to address the review question.[20] Only ITCs submitted by the sponsor are considered in the CDA-AMC review. Various statistical methods can be used to conduct ITCs, including NMA and population-adjusted methods. The validity of the results produced by ITC methods depends on how well the underlying assumptions of the particular method are fulfilled. Therefore, the decision to proceed with an ITC should be preceded by a feasibility assessment that provides support to the underlying assumptions of the method; all assumptions and their rationale should be reported in the sponsor submission.

The type of methods used for the submitted ITC should be those most likely to provide valid estimates given the review question and available evidence. The choice of ITC method should be based on the feasibility of a connected network, the evidence of heterogeneity between and within studies, the overall number of relevant studies, and the availability of individual patient data (IPD). For example, the Bucher method and NMAs provide suitable options when no IPD are available, whereas MAICs and simulated treatment comparisons are common techniques in the case of single-arm studies[21] (refer to Table 2 for more details). If multiple methods could be suitable, then the chosen method should be justified, considering the potential impact of that choice on the treatment effect estimates. If 1 or more ITCs are available in the published literature, the rationale for undertaking a new ITC (rather than relying on literature-based ITC[s]) or for selecting 1 or more ITCs for submission among multiple overlapping published ITCs should be justified. In such instances, the sponsor should aim to balance methodological quality (e.g., risk of bias), recency, and relevance.

When comparisons use IPD for the treatment and/or comparator in a population-adjusted comparison, the comparisons are potentially at greater risk of bias. Therefore, it is important that analysis methods are clearly specified in the protocol or statistical analysis plan and the rationale for why IPD were used should

be transparently reported in the sponsor submission. When IPD from RWD sources are used to create an external control for comparison with IPD from a single-arm trial, CDA-AMC would consider this to be RWE and would appraise such studies as per methodology for comparative RWE (e.g., the yet unpublished International Society for Pharmacoepidemiology RWE Appraisal tool).

Unadjusted (naive) ITCs that compare the results of individual arms from different trials as if they had come from the same RCT, or that compare the individual effects of 2 drugs from different trials with a common comparator, are not appropriate for drawing conclusions about the comparative effectiveness and harms of treatments and comparators and are discouraged. These comparisons have an increased chance for confounding factors to influence the results leading to biased and potentially misleading estimated treatment effects. If submitted, such ITCs are interpreted by CDA-AMC reviewers considering the known limitations and other evidence available within the sponsor's submission.

Methods for ITCs

Table 2 shows a summary of commonly used methods for ITCs, along with key considerations. However, ITC methods are a rapidly evolving field; therefore, Table 2 is not intended to be a comprehensive list of all possible methods.

Table 2: Summary of Commonly Used ITC Methods for HTA

ITC type	Type of trial data	Description and key considerations
Adjusted indirect comparison (Bucher method)	Aggregate trial data	• Simplest form of ITC • Estimates the relative treatment effect for a treatment in a simple network of common comparators • Preserves within-study randomization • Must meet assumptions of exchangeability (similarity and homogeneity), transitivity, and consistency: for the NMA to be valid, exchangeability should be plausible for every possible indirect comparison and consistency should be demonstrated for every direct and indirect comparison
Mixed treatment comparison NMA	Aggregate trial data	• Includes both direct evidence from head-to-head trials and indirect evidence derived through a common comparator treatment • Preserves within-study randomization • Must meet assumptions of exchangeability, transitivity, and consistency: for the NMA to be valid, exchangeability should be plausible for every possible indirect comparison and consistency should be demonstrated for every loop of evidence within the network
MAIC	Individual patient data and aggregate trial data	• Typically includes IPD for the drug product under review and aggregate trial data for the comparator(s) • Applies propensity score weighting to balance the study populations' baseline characteristics before performing indirect comparison • Anchored comparisons require adjustment for all effect modifiers • Unanchored comparisons require adjustment for all effect modifiers and all prognostic factors

ITC type	Type of trial data	Description and key considerations
STC	Individual patient data and aggregate trial data	• Allows comparison between treatment and comparator using IPD and aggregate data • Applies regression-based modelling of the relationship between baseline characteristics and outcomes • Regression model is fit to the trial with IPD for the outcome of interest and is then used to predict or simulate the expected outcome in the trial population with aggregate data only • Anchored comparisons require adjustment for all effect modifiers • Unanchored comparisons require adjustment for all effect modifiers and all prognostic factors
ML-NMR	Individual patient data and aggregate trial data	• Extension of the NMA framework to synthesize evidence from a mix of IPD and aggregate data across a network of trials comparing multiple treatments (2 or more studies) • Fits an individual-level regression model using IPD from at least 1 trial to allow the inclusion of patient-level prognostic factors and effect modifiers that may influence treatment effects • The individual-level model is integrated over the joint covariate distribution from trials where only aggregate data are available • Only applies to an anchored network • Allows for covariate-adjusted inferences from a network of trials for several comparators • Often limited by the lack of sufficient number of aggregate data trials for each treatment to meet the shared effect modifier assumption

IPD = individual patient data; MAIC = matching-adjusted indirect comparison; ML-NMR = multilevel network meta-regression; NMA = network meta-analysis; STC = simulated treatment comparison.

Studies Addressing Gaps in the Clinical Interventional Trial Evidence

The purpose for inclusion of any additional studies should be explicitly detailed in the sponsor submission. The only studies addressing gaps that are considered in the CDA-AMC review are the ones submitted by the sponsor (i.e., CDA-AMC does not search and extract other study data from publications or from other sources). Sponsors are encouraged to engage and consult with CDA-AMC as early as possible to help ensure reviewers understand how the supplementary evidence will help address critical gaps in the clinical trial data (i.e., additional evidence on selected populations [subgroups, or expansion beyond the population as described by the indication], other dosing regimens or treatment durations, effectiveness and harms, comparative effectiveness and harms, or long-term [comparative] effectiveness and harms, and so on).

Refer to the subsequent section on RWE for additional details.

Critical Appraisal of Pivotal and Other Clinical Interventional Trial Evidence

For each review, CDA-AMC considers which method(s) and tool(s) are most appropriate for the clinical evidence appraisal, with documentation of the choice(s) by the reviewer in the clinical report. In some cases, no relevant tool may exist, and reviewers may instead refer to methodological best practices.

For each submitted trial, internal trial validity is appraised for each relevant effect estimate by 1 CDA-AMC reviewer with independent verification.

External trial validity is appraised at the level of the body of evidence from the submitted clinical trials. Relevant tools may also be consulted for this appraisal, with explicit documentation in the clinical report.

To facilitate the appraisals, the sponsor-submitted evidence should be reported in accordance with relevant minimum reporting standards such as the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) statement[22] and the Consolidated Standards in Reporting Trials (CONSORT) guidance[23] and relevant extensions.[24] It is also expected that sponsors report the estimands of interest within their submission. Insufficient reporting may impact the ability of CDA-AMC reviewers to comprehensively appraise the sponsor-submitted evidence.

Internal Trial Validity

Central to the appraisal of internal trial validity is the assessment of the risk of bias, which is defined as "a systematic error, or deviation from the truth, in results."[25] Biases can result in underestimation or overestimation of true intervention effects. Since it is usually not possible to quantify the extent of bias with certainty,[7] only the risk of bias is discussed in the review; the direction of the potential bias is reported when it can be predicted.

The evidence appraisal includes specific aspects of study design and conduct that have empirically been observed to introduce bias.[26,27] Risk of bias is appraised at the level of the reported effect(s) of interest for each important outcome. Most commonly, the relevant effect is that of assignment to the intervention (i.e., the treatment policy estimand).

The specific domains of risk of bias that are appraised are unique to the design(s) of the study (or studies) under review. The most common designs included within clinical evidence submissions to CDA-AMC are parallel-group RCTs and single-arm trials submitted as pivotal trials to Health Canada. Relevant tools that may be consulted include the Cochrane Risk of Bias Tool version 2 (RoB 2)[26] and the National Institute for Health and Care Excellence (NICE) Methodology Checklist for RCTs.[28]

Domains of Bias Assessed in Parallel-Group RCTs[26]

- **Bias in the randomization process** includes an appraisal of the adequacy of the methods for generating and concealing the randomization sequence before the assignment of patients to the interventions. This also includes checking for imbalances in baseline demographic and disease characteristics across treatment groups as an indicator of the success of the randomization in achieving prognostic balance.
- **Bias due to deviation from the intended interventions** includes an appraisal of the adequacy of methods to blind participants, caregivers, and personnel to the assigned treatment following randomization. Potential inadequacies include open-label trials, or when it is likely that patients and/or outcome assessors became unblinded (e.g., due to treatment-related specific AEs). In those cases, this domain also includes an appraisal of whether deviations from the intended interventions (e.g., failure to implement the interventions as intended, lack of adherence, treatment crossover, or implementation of nonprotocol interventions such as concomitant treatments) occurred due to trial context.

- **Bias due to missing outcome data** includes an appraisal of whether complete outcome data are available for all, or nearly all randomized participants, and if not, whether missing outcome data are differential across distributions of baseline covariates, and/or at risk of inducing confounding. When outcome data are missing, this includes an appraisal of whether the missingness depends on the true value of the outcome and certain measured or unmeasured patient characteristics (i.e., missing at random or missing not at random). When available, sensitivity analyses are examined to appraise whether the results are sensitive to the assumptions of the missing data imputation based on statistical model.
- **Bias in measurement of the outcome** includes an appraisal of the appropriateness of the method or tool used to measure the outcome (refer to "Validity of Outcome Measures"). Particular attention is paid to differential misclassification of the outcome, measurement error, and the adequacy of methods to blind outcome assessors, who may be study participants in the case of patient-reported outcomes, to the assigned treatment. Where inadequacy in blinding exists (e.g., in open-label trials or when it is likely that patients and/or outcome assessors became unblinded due to treatment-related specific AEs), this includes an appraisal of how the outcome may have been influenced by knowledge of the treatment assignment (degree of subjectivity).
- **Bias in selection of the reported result** includes an appraisal of whether the available effect estimates are the result of analyses that were prespecified in the study protocol and/or statistical analysis plan before unblinding of outcome data, whether the numerical results presented are likely to have been chosen from multiple available outcome measurements or analyses of the data (e.g., due to a favourable magnitude or direction of effect).

Domains of Bias Assessed in Other Trial Designs

Additional domains are appraised for other pivotal trial designs that did not use randomization to allocate patients to comparison groups. Relevant tools that may be consulted include those intended for the appraisal of nonrandomized studies, including the Cochrane Risk of Bias in Nonrandomized Studies of Interventions (ROBINS-I) tool.[27] Domains of bias include, but are not limited to, the following.[27]

- **Bias due to confounding** includes an appraisal of the potential for baseline or time-varying measured or unmeasured confounding between the intervention and the outcomes of interest; whether there was the ability to measure all relevant confounding variables, and whether the measurement was valid and reliable; and whether appropriate study design and/or analysis methods were used to control for all important confounding domains. Unmeasured confounding can also be assessed (e.g., through negative controls and other quantitative bias assessment methods).
- **Bias in selection of participants** includes an appraisal of whether selection into the study was based on characteristics observed after the start of the intervention, whether follow-up and the start of the intervention correspond for most participants (i.e., inception or lead-time bias), and/or whether interventions are defined in such a way that there is a period in which the outcome cannot occur (i.e., immortal time bias), and whether appropriate analysis methods were used to correct for potential selection bias.

- **Bias in classification of the interventions** includes an appraisal of whether intervention status was classified correctly for all or most participants, whether the information used to define intervention groups was collected at the start of the intervention, and whether there is the potential for differential misclassification of the interventions between groups (e.g., due to knowledge of the outcome or risk of the outcome).

Considerations for Single-Arm Trials and LTE Studies

Generally, causal interpretations cannot be drawn from single-arm trials.[29] Due to the lack of randomized comparator, observed effects may be attributable to the effect of the drug product, placebo effects, or the natural history of the disease. When drawing conclusions from single-arm pivotal trials, evidence reviewers consider internal trial validity as well as information external to the trial (e.g., natural history, external control estimates, clinical expert input) to estimate the benefits and harms of the drug product relative to appropriate comparator(s).

A European Medicines Agency reflection paper has outlined potential sources of bias that can be considered for single-arm designs,[29] some of which apply to open-label RCTs, and others apply to RWE or observational studies. LTE studies, which are often single-arm in design, are appraised using the same methodological approaches as those used for single-arm trials submitted as pivotal evidence.

Other Considerations

Other critical appraisal points may be considered depending on the individual study design and circumstances. In addition to domains of bias, the adequacy of control for multiple comparisons (risk of type I error or erroneously rejecting the null hypothesis) across end points and interim analyses is appraised. End points not controlled for multiplicity are considered to provide supportive evidence. Other items that may be appraised include but are not limited to issues such as the study's power to detect a statistically significant effect, inconsistency in effect estimates across studies, the precision of the effect estimates, the adequacy of the length of follow-up, the plausibility of assumptions underlying statistical models, the appropriateness of noninferiority margins, the impact of interim analyses and early stopping, and potential limitations of specific end point types (e.g., composite or surrogate end points).

Validity of Outcome Measures

Reviewers appraise the validity of outcome measurement instruments used within studies included in the submission, based on evidence supporting validity submitted by the sponsor and considering relevant guidance documents.[30-32] This may also be supplemented with clinical expert opinion or other relevant literature as needed. This includes an assessment by the CDA-AMC review team of the relevance of the selected instrument(s), and the following properties[33]:

- **content validity:** the degree to which the content of the instrument reflects the concept that it is intended to measure
- **construct validity:** the adequacy with which scores of the instrument are consistent with hypotheses (e.g., relationships with scores on other instruments, differences across relevant groups, relationships with other outcomes) about the concept that it is intended to measure

- **criterion validity:** the degree to which scores of the instrument are consistent with a reference standard (e.g., comparing abbreviated to full versions of an instrument)
- **reliability:** the degree to which error across a series of measurements is minimized, such that scores for the same measure, in the same individuals, would be similar when measured repeatedly; this includes scores using a different set of items from within the same tool (internal consistency), over time by the same rater (test-retest), by different raters at 1 time (interrater), or by the same rater at different times (intrarater)
- **responsiveness:** the degree to which an instrument can detect changes in the concept of interest over time.

Additionally, when the sponsor refers to minimal important differences (MIDs) within the submitted evidence, reviewers appraise the quality of the evidence supporting the MID for the outcome in the target population, considering methodological best practices as appropriate,[34,35] and drawing on the published literature as needed. CDA-AMC does not typically consider minimal clinically important differences, unless it has been explicitly established for the specific end point and accepted as clinically meaningful by patients and/or clinicians. The strength of evidence supporting the validity of a surrogate end point is also appraised (refer to the section on Outcomes: Surrogate Outcomes).[36,37]

External Trial Validity

External trial validity is sometimes termed generalizability, applicability, or directness. Important factors that may influence the generalizability of the findings to patients living in Canada should be defined for each PICO(T)(S) element. Evidence that is considered generalizable comes from studies that enrol patients who are reflective of those seen in clinical practice in Canada; for example, studies that measure outcomes relevant to patients, clinicians, and public drug plans, are assessed at clinically relevant follow-up times, and occur in clinical practice settings relevant to Canada, as with pragmatic trials.

Interpretation of Subgroup Effects

Relevant subgroup effects are presented and appraised for the primary end points of included trials. Subgroup analyses for other end points may be presented and appraised when relevant to decision-making. CDA-AMC reviewers consider that any subgroup analyses (within or across trials) submitted by sponsors are typically used to assess consistency of treatment effects across groups of patients. Subgroup analyses may be used to explore subgroup differences, identify a subgroup of the population where the benefit is more evident, within a trial where evidence of benefit to the full trial population is equivocal or unconvincing.[38,39]

Ideally, the results of subgroup analyses are presented visually on forest plots. Commonly, CDA-AMC reviewers visually inspect point estimates and confidence intervals across various subgroups for consistency in direction and magnitude. When effects appear consistent, this might be considered to strengthen the applicability of the results across the full trial population.[39]

In some cases, the indication and/or reimbursement request may be for a single subgroup of the full population within included trials. In these cases, the focus of the appraisal is on the single subgroup alone, not all subgroups.

If there is evidence of potential effect modification (i.e., there is observed variability in estimated effects by patient or disease characteristics) believed to be potentially relevant to decision-making, reviewers may assess the credibility of the effect modification using appropriate tools (e.g., Instrument to Assess the Credibility of Effect Modification Analyses [ICEMAN]),[40] which would be cited in the CDA-AMC clinical report. Considerations include the following[40]:

- whether the direction of the effect modification was hypothesized a priori
- whether the effect modification was supported by prior evidence
- whether a test for interaction suggested that chance was an unlikely explanation for the apparent effect modification
- whether the investigators tested only a small number of effect modifiers or considered the number (e.g., via multiplicity control) in their statistical analysis
- for continuous variables, whether arbitrary cut points were avoided
- any other considerations that may increase or decrease the credibility, on a case-by-case basis.

Interim Results

In certain situations, interim results are submitted for appraisal by the CDA-AMC review team. The potential for overestimation of the treatment effect should be noted for interim study results, thus the adequacy of adjustments for multiplicity are appraised by the CDA-AMC reviewer, in addition to other considerations listed in this methods guide that apply to all HTA appraisals of final results.

Certainty of Evidence Assessment

Framework and Domains Considered

The certainty in the body of clinical trial evidence (which may be composed of 1 or more pivotal trials or other RCTs) for each selected effectiveness or harm comparison outcome is assessed according to GRADE guidance.[19] The certainty of evidence from RCTs begins at high and is rated down (to moderate, low, or very low) for uncertainty related to the following factors:[41]

- study limitations (risk of bias)
- inconsistency in effects across studies
- indirectness
- imprecision in effects
- publication bias.

The certainty of evidence from bodies of evidence composed of pivotal studies that did not use randomization to allocate patients to comparison groups (including single arms trials, with or without a comparison to an external control estimate) is generally started at low (acknowledging the likely risk of selection bias and residual confounding), with the opportunity to rate up (to moderate or high) when, in the absence of other serious limitations:[41]

- there is a large magnitude of effect

- there is a dose-response gradient
- all plausible residual confounders or biases would reduce an apparent treatment effect.

The certainty of evidence for outcomes assessed in single-arm trials without any formal external comparison is generally started at very low without the opportunity to rate up, acknowledging that this study design does not allow for definitive conclusions regarding the efficacy or harms of a drug product relative to any comparator.

For each relevant outcome, the CDA-AMC review team prioritizes rating certainty in a clinically important effect, based on information submitted by the sponsor (e.g., MIDs when provided by the sponsor, literature, and/or consultation with clinical expert[s]). When no threshold for clinically important (relative) treatment effect can be determined, the review team may assess certainty in any effect (i.e., non-null effect). In such cases, the clinical importance of the estimated between-group differences is uncertain and is explicitly described as such.

Summary of Findings Tables

Results of the certainty of CDA-AMC evidence appraisals are reported in Summary of Findings tables,[42] including footnotes to transparently detail the reasons for rating the certainty of evidence down (or in rare cases, rating up).[19]

Critical Appraisal of ITCs

ITCs are increasingly used to evaluate the comparative effectiveness and harms of treatments and comparators in the absence of direct head-to-head trials. However, ITCs make assumptions that, if violated, can lead to biased results. Few validated ITC appraisal tools and checklists exist, but CDA-AMC reviewers may use appropriate tools to aid in their appraisals with the use of any such tools explicitly documented in the clinical report. Other HTA organizations have technical guidance for the conduct and reporting of ITCs that may also be helpful to CDA-AMC reviewers for critically appraising sponsor-submitted ITCs. For example, technical guidance documents are available from the National Institute for Health and Care Excellence (NICE) Decision Support Unit,[43-47] as well as from the International Society for Pharmacoeconomics and Outcomes Research.[20,48] When applicable, reviewers will document the choice of tool(s) applied in the ITC appraisal in the CDA-AMC clinical report.

To facilitate the appraisals, the sponsor-submitted evidence should be reported in accordance with relevant minimum reporting standards, such as the PRISMA extension for NMAs (PRISMA-NMA),[49] with consideration for the updated guidance in PRISMA 2020 (*PRISMA Extension for Reviews Incorporating NMA*).[18] While there is a GRADE approach to the assessment of the certainty of NMA estimates, this application of GRADE is not used by CDA-AMC reviewers at this time, as this assessment should be completed by those conducting the evidence synthesis (of the SR and NMA). The detailed information required to inform such an assessment, such as risk of bias appraisals for each estimated effect, direct and indirect effect estimates, and absolute effect estimates based on assumed baseline risk, are not typically provided by the sponsors which limits the reviewer's ability to apply GRADE. Sponsors are encouraged to provide a thoroughly

detailed and transparent technical report of the evidence base and assumptions underlying the ITC results to improve the interpretability and credibility of the ITC findings.

Feasibility Assessment

An important part of the critical appraisal is identifying whether a submitted ITC was feasible. It is a best practice to assess the feasibility of conducting an ITC before starting the analysis. This involves evaluating whether there is enough high-quality data and appropriate conditions to ensure the comparison will be valid and reliable. The sponsor's submitted ITC technical report should document the feasibility assessment and support the decision to proceed with the ITC. If the methods and results of the feasibility assessment are not reported or a feasibility assessment was not conducted, then this will be indicated in the CDA-AMC clinical report.

In brief, the feasibility assessment should include the following elements.

- **Evidence synthesis:** This takes the form of an ITC based on a systematic literature review to identify all relevant studies and treatment comparisons following the PICO(T)(S) framework. Sponsors may leverage the SR to identify pivotal and other trial evidence for their submission. The process and rationale for identifying, selecting, and including or excluding relevant RCTs should be provided in the technical report.
- **Network structure evaluation:** This is a network diagram to visualize direct and indirect comparisons among treatments, to allow the determination of whether methods for connected (anchored) or unconnected (unanchored) networks are feasible.
- **Determination of availability of IPD:** This involves the identification of trials that have IPD available to inform the use of population-adjustment methods, and to assess the quality and completeness of IPD for capturing key effect modifiers and prognostic factors.
- **Check of similarity, homogeneity, and consistency**[63]:
    - **Similarity:** This involves the assessment of distributions of potential effect modifiers across trials to judge plausibility of the constant relative effects assumption (similarity) required for NMAs. A comprehensive list of potential effect modifiers should be established before conducting the evidence synthesis. This list should be informed at least by findings from prior studies on the therapeutic indication and by clinical expert input. The process for identifying relevant effect modifiers and the rationale for which ones were considered to apply to the comparisons should be transparently documented. Numerous qualitative and quantitative methods exist for assessing similarity. These range from visually comparing study designs and PICO characteristics to meta-regression analysis to assess the impact of study-level covariates on treatment effects between trials. Combining descriptive and statistical approaches is more likely to detect similarities and differences between studies than relying on a single method. The process should be systematic, thorough, and transparently documented.
    - **Homogeneity:** This can be assessed qualitatively or quantitatively. Qualitative homogeneity (also referred to as exchangeability) exists if each trial estimates the same single treatment effect or different treatment effects distributed around a typical value. Quantitative homogeneity is tested

using statistical methods like the Q-test and the I^2 heterogeneity measure. If statistical tests for homogeneity are conducted and reported, conclusions regarding homogeneity should also be informed by descriptive or qualitative comparisons.

- **Consistency:** This refers to the agreement between the direct and indirect estimates of the treatment effects, that is, that the direct and indirect evidence sources estimate the exact same parameters. Consistency relies on the assumption of exchangeability: if this assumption holds then the direct and indirect estimates should be similar.

- **Study quality:** Risk of bias assessment of relevant effect estimates in each trial should be done before the analysis and may impact the makeup of the base-case model. The included studies should be of similar quality and free from systematic biases that could affect the comparative treatment effect estimates. High variability in study quality can violate the exchangeability assumption. The rationale for why a trial with a high risk of bias, for example, was included in (or excluded from) the base-case model should be reported and sensitivity analyses performed to explore the impacts on the results from the decision.

The feasibility assessment should clearly describe how these factors were used to determine if the available evidence network can produce comparisons relevant to the target populations and decision context.

ITC Limitations

Table 3 describes, at a high level, considerations in the assessment of the sponsor-submitted ITC(s). The feasibility assessment for conducting an ITC, if provided, will help inform this appraisal.

Table 3: Considerations for the Assessment of ITCs

Area of consideration	Topic	Sample guiding questions or principles
General considerations	Rationale for use of ITC	• Why is the ITC required? Justification for use of the ITC should be provided given that ITCs have the potential for less certainty relative to direct evidence due to the assumptions that must be met. • What does the ITC add if direct head-to-head evidence vs. relevant comparators is available?
	Research question	• What is the research question that the ITC addresses? Is the estimand adequately defined? • Is the research question relevant to the review objectives? If not, are deviations adequately justified? • Was the target population specified in the research question described? Details about how the target population relates to the Health Canada indication, reimbursement request, or other subpopulations (e.g., by specific lines of treatment) should be included.
Assessing the evidence base and network	Systematic literature review	• Was there a predefined protocol? Are deviations from the protocol adequately justified? • Were inclusion or exclusion criteria per the PICO(T)(S) framework clearly defined? Are the PICO(T)(S) relevant? • Was there a comprehensive literature search?

Area of consideration	Topic	Sample guiding questions or principles
		• Do methods for study selection and data extraction minimize error and bias? • Was the study selection, data extraction, and risk of bias assessment performed independently by 2 reviewers or done by 1 reviewer and checked by another? • Were included studies described in adequate detail, and is justification provided for excluded studies?
	Network geometry	• Is there a connected network of trials linking the treatment(s) of interest through a common comparator(s)? • In a connected network, is the rationale for the choice of common comparator(s) provided? • Are there closed loops (direct evidence)? • Are the number of studies contributing to each comparison provided?
	Assumptions (similarity, homogeneity, consistency)	• Is there sufficient similarity across trials in terms of study design, populations, interventions, and outcomes? • Are the distributions of potential effect modifiers sufficiently balanced across trials in the network? • Is there evidence of heterogeneity for pairwise comparisons? • Was the quality of the individual trials assessed and reported? Comment on the approach used and its appropriateness to interpret trials with a high risk of bias. • Is there a risk of inconsistency between direct and indirect evidence?
Assessing the analysis methods	Network and adjustments	• If a connected network exists, were standard NMA methods used when assumptions appeared plausible? • If population adjustment was used, is clear justification provided for why this was needed (e.g., nonconnected network)? • Are the assumptions underpinning the anchored vs. unanchored comparison(s) described and met?
	Statistical methods	• Were the statistical procedures transparently described in sufficient detail? • Was appropriate rationale provided for the choice of statistical procedures? • Do methods for study selection and data extraction minimize error and bias? • Was a satisfactory method used to appraise risk of bias for each relevant effect estimate? • Tools that may be leveraged in the appraisal of ITCs include the AMSTAR 2[50] tool to assess quality and ROBIS[51] tool to assess risk of bias.

AMSTAR 2 = A Measurement Tool to Assess Systematic Reviews 2; ITC = indirect treatment comparison; NMA = network meta-analysis; PICO(T)(S) = population(s), intervention(s), comparator(s), outcome(s), time or time frame, study design or setting; ROBIS = Risk of Bias in Systematic Reviews; vs. = versus.

Real-World Evidence

Real-World Data[52]

RWD are data relating to patient status and/or the delivery of health care collected from a variety of sources, and can include electronic medical records, clinical and disease registries, and administrative databases.

Real-World Evidence[52]

RWE is evidence on the use, safety, effectiveness, and cost of health technologies that is derived from RWD.

Prospectively planned RCTs continue to be the most robust study design for estimating the causal effects of interventions. However, the generalizability of RCTs is often limited, and RCTs may not address all research questions relevant to assessments of comparative clinical effectiveness and harms. Additionally, the conduct of RCTs is not always feasible in certain diseases or disorders (such as rare diseases, because of the limited number of patients), or for ethical reasons (in populations such as children, patients who are pregnant, or older adults).

Best practices in using RWE includes the need to integrate it with other evidence sources.

Examples of situations where additional RWE studies may inform the evidence submission include:

- primary evidence for situations in which it is not feasible or ethical to conduct a robust RCT that will address the HTA question of comparative effectiveness and harms
- to address evidence gaps and/or remaining uncertainties regarding comparative effectiveness and harms
- to provide evidence of long-term safety and effectiveness (durability of treatment effect)
- additional evidence on selected populations (subgroups, or expand beyond the population as described by the indication), or on other dosing regimens or treatment durations
- to supplement clinical trial data by providing additional data on patient-reported outcome measures and patient-reported experience measures related to patient preferences, values, health, experiences, or goals for care and treatments, collected outside of a controlled clinical trial
- to provide contextual information (note that the types of evidence outlined subsequently are not generally part of the clinical evidence submitted by the sponsor to address the research question of comparative clinical effectiveness or harms, or long-term [comparative] effectiveness; these evidence types may provide input to the health economic analyses, or other information to contextualize decision-making):
    - the use of medication in the real world (e.g., duration of treatment, persistence, adherence)

- burden of illness studies to characterize health conditions and patient populations (as defined by the indication, or a subpopulation), and/or to understand the current treatment setting (local standard of care treatments, care pathway from diagnosis through treatments, and so on)
- input for economic models: incidence and prevalence, baseline rates of events transition probabilities, health care resource utilization, and so on.

Reporting of RWE

The CDA-AMC *Guidance for Reporting Real-World Evidence*[52] provides guidance on the clear, complete, and transparent reporting of study methodology and findings, as outlined in Table 4. This fosters credibility and trust in the results and facilitates appraisal by CDA-AMC reviewers.

Table 4: Summary of the CDA-AMC *Guidance for Reporting Real-World Evidence*[52]

Section	Key considerations
Section 1: Study design and research questions	Report a clearly stated aim and study questions and any specific sub-questions.
	Report the study design.
	Include the study protocol developed a priori.
Section 2: Setting and context	Describe type of care setting and geographical location or other information to contextualize the data source; justify transportability of data from across jurisdictions within Canada, or from outside of Canada, if used.
	Describe all relevant study period dates, including periods of recruitment, exposure, follow-up, and data collection.
Section 3: Data specifications	Describe data source, data collection methods, data quality, and if any linkage was performed.
Section 4: Data sources	Justify that the data source contains appropriate elements to address the question(s).
	Describe how variables of interest were assessed; provide or reference any data dictionary.
Section 5: Participants	Describe who is in the study and how they were identified; provide inclusion and exclusion criteria, patient selection at each stage, and so forth.
Section 6: Exposure definitions and comparators	Define requirements for exposure definition (e.g., single, multiple, or continuous) and relevant start and stop windows.
	Define and justify any comparator(s).
Section 7: Outcomes	Report definitions for all study outcomes (primary, secondary, and exploratory) and results of any validation studies supporting their use.
Section 8: Bias and confounding	Describe procedures used to address potential sources of bias.
	Describe how confounder variables were selected, their distribution among groups, and potential for unmeasured confounders or effect modifiers.
Section 9: Statistical methods	Describe the statistical methods (i.e., justification for selected methods, variable selection for models, missing data, handling of follow-up time, propensity score estimation, approaches for handling intercurrent events using the estimand framework, and so forth).
Section 10: Study findings	Specify the number of patients included and excluded for each analysis and describe the characteristics of the study population, including of exposure groups.

Section	Key considerations
	Report all unadjusted and adjusted estimates, including measures of precision, for all prespecified primary and secondary analyses.
	Report all other prespecified analyses (e.g., subgroup or sensitivity); avoid selective reporting of results.
Section 11: Interpretation and generalizability	Interpret the findings, including precision and if findings are clinically meaningful.
	Discuss the validity of the results within the sample and their relevance in the target population, including transportability of findings across jurisdictions within Canada, or to Canada if using international data.
	Describe if results are consistent with prior known information and if not, provide an adequate explanation.
Section 12: Limitations	Provide a consideration of limitations of the study, including the impact of bias, potential confounding, and assumptions.
Conflicts	Report conflicts of interest: role of sponsor and source of funding.

CDA-AMC = Canada's Drug Agency.

Appraisal of RWE

All RWE is appraised generally as per usual methodology for appraisal of observational studies. Appraisal of the evidence from real-world studies of comparative effectiveness, including clinical trials that use RWD to form an external control arm, may be facilitated by the use of tools such as the yet unpublished International Society for Pharmacoepidemiology RWE Appraisal tool, and the ROBINS-I tool,[27] or other tools as deemed appropriate by the CDA-AMC reviewer. The application of a selected appraisal tool would be documented in the clinical report.

Appraisal domains considered include data source and quality, study design, data analysis, confounding, and bias.

Data source and quality: RWE uses various data sources, including but not limited to electronic health records, health care administrative databases, patient registries, and pharmacy and lab data. CDA-AMC reviewers consider the relevance of the data source(s) and the suitability of the data elements within the database to generate information that would address the research question of the RWE study; the sponsor should provide sufficient detail to facilitate this assessment. Incomplete, inaccurate, or missing data can lead to flawed analyses and unreliable conclusions.[53] Relevant tools, such as the International Society for Pharmacoeconomics and Outcomes Research SUITABILITY checklist for assessing RWD from electronic health records,[54] may be used by CDA-AMC reviewers in the appraisal of data quality for relevant comparative RWE studies.

Study design: High-quality nonrandomized studies can produce valid estimates of relative treatment effects in certain situations, provided appropriate study design and statistical methods are used to minimize and control for bias and confounding. Common study design principles to improve that likelihood include target trial emulation, inclusion of new users of the drug product rather than those who have been using the drug product for some time (prevalent users), the use of active comparators or alternate interventions for the

same indication, negative controls (nonusers), and target or final clinical outcomes such as mortality. These study design elements are best depicted graphically,[55] which facilitates the appraisal of the appropriateness of the study design to address the research question by the CDA-AMC reviewer.

Data analysis, bias, and confounding: CDA-AMC reviewers consider the appropriateness of the statistical analysis, including the types of methods chosen and whether all known confounding variables are appropriately controlled. Biases can result in underestimation or overestimation of true intervention effects. RWE has greater vulnerability to biases, hence the assessment of a lower certainty in the treatment effect relative to other relevant comparators or different care options. Domains of bias assessed by the International Society for Pharmacoepidemiology's RWE appraisal tool for comparative RWE studies include: the potential for bias due to exposure or outcome misclassification (misclassification bias), study design biases due to study design decisions (e.g., immortal time bias), the potential for bias due to residual confounding, the potential for bias due to suboptimal implementation of propensity scores, and the potential for bias due to missing and suboptimal handling of missing data. The 7 domains of bias assessed in the ROBINS-I assessment tool are: confounding, selection bias, bias in measurement classification of interventions, bias due to deviations from intended interventions (performance bias), bias due to missing data, bias in the measurement of outcomes, and selective reporting bias.[27] Refer also to the earlier section describing common risk of bias assessments in other trial designs.

Any scenario or sensitivity analyses undertaken should be detailed in the sponsor submission, as these may support the estimation of the relative treatment effect and mitigate uncertainty.

Output of the Clinical Evidence Review

Informative Statements

Informative statements on the CDA-AMC appraisal of the sponsor-submitted clinical evidence are intended to inform the committee in their deliberation. Both GRADE and other informative statements are intended to support the committee in efficient deliberations by explicitly summarizing the level of certainty of any statements about the clinical evidence in a consistent and transparent manner.

GRADE Informative Statements

GRADE informative statements[56] are used to describe the certainty of evidence included in the SR (pivotal trials and other RCTs) for each important outcome, both in the Summary of Findings tables and in the CDA-AMC reviewer report text. The statements describe the magnitude, direction, and certainty of the observed effects, as follows:

- results in high-certainty evidence
- likely results in moderate-certainty evidence
- may result in low-certainty evidence
- very uncertain, very low-certainty evidence.

Other Informative Statements

CDA-AMC reviewers will use informative statements to facilitate clear and transparent reporting of the findings of the CDA-AMC evidence appraisal. Some examples of potential informative statements are shown here.

> Based on the (direct or indirect) comparison, patients (as per indication) treated with (the intervention) (description of change [e.g., improved or had similar effect, or other appropriate terminology]; the efficacy comparison end point[s]); compared to (relevant comparator[s]). Relative safety and tolerability were (description of change [e.g., improved or similar to]) (relevant comparator[s]).
>
> (This would be followed by the [summary of] evidence to support the informative statement.) For example, where drug X is the drug product under review:
>
> Based on direct comparison, patients with disease or condition ABC treated with drug X demonstrated improved overall survival (OS) compared to drug Y. Safety and tolerability were similar for drug X compared to drug Y.
>
> Evidence: The previous statements are based on results from Study 123: RCT of X versus Y in (patient population), with primary end point of OS.

An alternative example might be as follows:

> In a MAIC including multiple studies, patients with condition ABC (as per approved indication) treated with drug X had improved OS compared to drug Y, and similar OS compared to drug Z. Harms were similar across drug treatments based on the rate of SAEs including death, common AEs, and rates of discontinuation of treatment.
>
> Evidence: The previous statements are based on a MAIC (summary of supporting evidence).

If RWE is submitted to address evidence gaps of (comparative) effectiveness and safety, or long-term benefits and harms:

- list the uncertainties (from clinical interventional trials)
- state if the RWE was designed to address the research question(s) that would address the evidence gap(s)
- provide conclusions from the RWE assessment.

For example:

- Uncertainties were efficacy in the population beyond that defined in trials, in particular patients with efficacy versus relevant active comparators (Y) or (Z), and duration of response beyond (number) weeks of treatment; that is, long-term effectiveness and safety.
  - Regarding the **efficacy in the population beyond that defined in trials:** The sponsor submitted 1 RWE study to provide additional evidence in the population with (condition). While these studies do add to the evidence for a positive effect on outcome (defined) as assessed by (outcome measure) in (patient population), this study did not address the evidence gap cited for the patient

population defined in the pivotal clinical trials (references), and the outcome (defined) in those trials, as assessed by (outcome measure).

- No new direct evidence was submitted to address the evidence gap of **efficacy or effectiveness** (based on improvement of [outcome] as assessed by [outcome measure]) **versus active comparators Y or Z, or long-term effectiveness and safety.**
- The new ITC did not overcome the limitations of the original ITC; uncertainties noted in the appraisal of the original ITC were not addressed by this new ITC. Hence, there is **no greater certainty in the comparative effectiveness and safety of drug X versus the other relevant comparators of Y and Z**, based on this new ITC.

Evaluation of Economic Evidence

Refer to *Guidelines for the Economic Evaluation of Health Technologies: Canada* for more information on this topic.[7]

Value Considerations Contributing to Decision-Making

The framework for deliberation and the resulting committee reimbursement recommendation is not only based on clinical or economic evidence, but also other value considerations, including but not limited to unmet needs, patients' experiences and values, clinicians' perspectives and preferences, perspectives of health care systems (including clinical experts and public drug plans), social and ethical considerations, health equity, and other relevant factors.

HTA aims to determine the value of a health technology at different points in its life cycle. Alongside clinical and economic value, other relevant dimensions of value include ethical, social, and organizational considerations, alongside wider implications for patients, caregivers, and other populations.[57] These elements of value may vary depending on the perspective taken, the parties involved, and the decision context. Though these are not the focus of this methods guide, elements of these are detailed in what follows.

Qualitative Research

Qualitative research can explore areas such as patient experiences, values, preferences, acceptability, equity, and implementation implications. Qualitative evidence can contribute an understanding of:

- the experiences and perspectives of patient subpopulations who have a disease or condition and are receiving (or not receiving) a treatment or test; this can also include caregiver and clinician perspectives
- the quality of life and experiences of patients with a disease or condition; this can also include caregiver and clinician perspectives

- the impact of treatments, tests, and health systems on the quality of life and experiences of patients with a disease or condition; this can also include caregiver and clinician perspectives
- the acceptability of different types of treatment or tests from a patient's perspective
- inequities or particular challenges in accessing treatments from a patient, caregiver, and clinician perspective
- the feasibility of implementing a health intervention from a clinician, health system, and societal perspective.

Data may be collected through formal qualitative research studies or from a systematic review of relevant qualitative research. Methods of analyzing, synthesizing, and presenting qualitative evidence include rapid review, framework synthesis, narrative summary and synthesis, meta-synthesis, and thematic synthesis.

Ethical Considerations

For certain CDA-AMC drug reviews deemed to be of higher complexity, or when particularly salient ethical considerations may arise, a dedicated Ethics Review will also accompany the Clinical and Economic Evidence Reviews, as per the CDA-AMC procedures.[60]

The objective of the Ethics Review is to identify and describe ethical considerations associated with the use of the drug product under review for its indicated purpose, including considerations related to the disease context, evidentiary basis, the use of the drug product, and impact on health systems.

The Ethics Review addresses several research questions, including but not limited to:

- What ethical considerations arise in the context of the indicated disease or condition, including considerations related to diagnosis, treatment, and outcomes?
- What ethical considerations arise in relation to the evidence (e.g., clinical and economic data) used to evaluate the drug product under review?
- What ethical considerations arise in relation to the use of the drug product under review for patients, their caregivers, and their clinicians?
- What are the ethical considerations for health systems related to the drug product under review?

Overview of Methods for Ethics Reviews

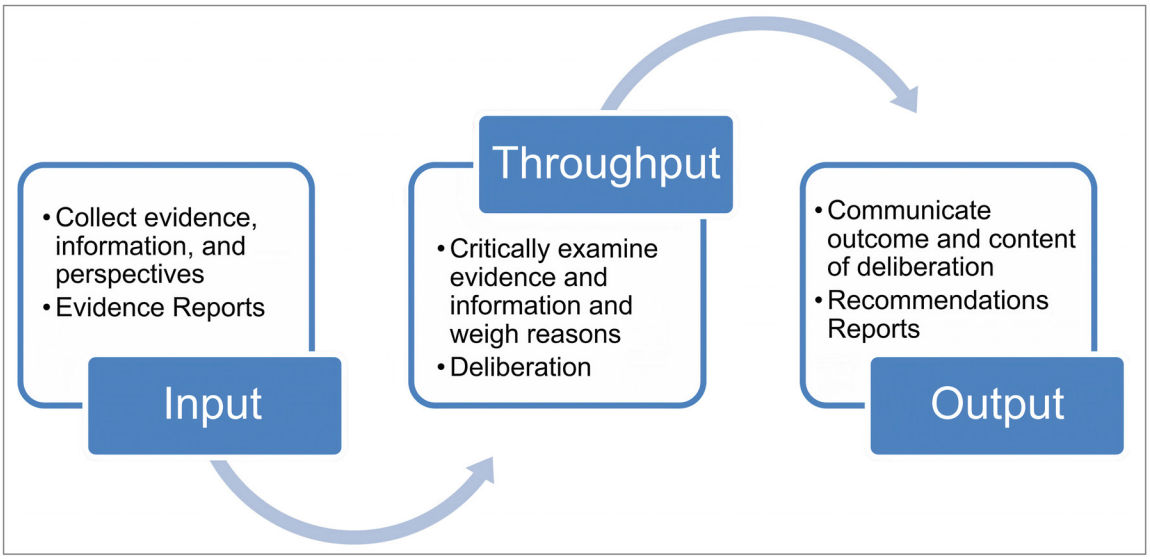
Guiding questions identified in the EUnetHTA Core Model 3.0 Ethics Analysis Domain,[3] and supplemented by relevant questions from the Equity Checklist for Health Technology Assessments (ECHTA),[58] drive the identification of ethical and equity considerations relevant to the use of the drug product under review in the treatment of the relevant disease or condition. These guiding questions are organized and analyzed to respond to the research questions.

The data used to inform ethics reviews draws on patient and clinician groups, clinical expert, and public drug program input collected during the CDA-AMC Reimbursement Review and a complementary search of the published literature.

Deliberation

During the expert committee meetings, committee members review CDA-AMC Evidence Reports and supporting materials (products of the input phase, described in the following figure), which assess the evidence, information, and perspectives relevant to the drug product under review. Following deliberation (the throughput phase), the committees issue recommendations or guidance (the output phase) in the form of published recommendation reports, which include a plain language summary.

Figure 2: Input, Throughput, and Output



Adapted from: Bond K, et al.[59] Copyright © 2020, Creative Commons CC BY https://creativecommons.org/licenses/by/4.0/. Adaptations made to simplify descriptions of each stage of the model.

During the expert committee meetings, committee members review CDA-AMC Evidence Reports and supporting materials from interested parties (products of the input phase) that assess the evidence, information, and perspectives relevant to the health technology under review. Following deliberation (throughput phase), the committees issue recommendations or guidance (output phase).

The guiding principles for deliberative processes reflect the overarching goals of the health systems that our recommendations are intended to support:

- **Need:** Allocating health care resources according to the severity and urgency of health conditions, capacity to benefit, and the acceptability, availability, and effectiveness of alternative health technologies.
- **Patient benefit:** Prioritizing health technologies that deliver net positive outcomes and improvements for individual or population health.
- **Health system sustainability:** Meeting the health and health care needs of the population in a way that leads to optimal health in the present without compromising availability of resources to current and future generations.

- **Health equity:** Distributing health care resources and arranging health care practices and systems to minimize unfair or avoidable disparities in health outcomes and experiences of care across the population.

These guiding principles are operationalized in the deliberation using a deliberative framework.

In evaluating health technologies, the expert committees are asked to consider 5 domains of value (Table 5):

- clinical value
- unmet clinical need
- distinct social and ethical considerations
- economic considerations
- impacts on health systems.

For more information on the deliberative framework and the CDA-AMC Expert Committee deliberations, refer to Expert Committee Deliberation at Canada's Drug Agency.[64]

Table 5: Summary of Deliberative Framework Domains

Domain	Description
Clinical value	The value that patients derive from a health technology in terms of its effect on their health and health-related quality of life. The determination of the clinical value of a health technology requires the measurement of its clinical benefits and harms and an assessment of the impact of these effects on patients. Clinical benefits and harms are assessed against relevant comparators.
Unmet clinical need	Morbidity and/or mortality arising from a condition or symptom that is not addressed effectively by available treatments.
Distinct social and ethical considerations	The social and ethical implications of health technologies not already assessed in other domains and how they affect patients, caregivers, populations, and the organization of health systems. It includes nonclinical needs, which are the social, psychological, and logistical factors that influence the appropriateness, accessibility, and acceptability of a health technology beyond its direct clinical outcomes. It also examines the broader social and ethical considerations related to the design, evaluation, and implementation of health technologies.
Economic considerations	Economic evidence to inform the financial, human, or other resource implications associated with the technology under review, and whether it is worthwhile to allocate resources to the technology under review given its expected clinical benefits. Considerations may include the potential resource or cost impacts of the technology under review versus relevant comparator(s).
Impacts on health systems	Two distinct but interrelated components: organizational feasibility of adoption is the ease with which the health technology can be implemented in the health system while realizing its clinical value, while economic feasibility of adoption examines how the adoption of a health technology will economically impact the payer or budget holder.

Process Elements

The Procedures for Reimbursement Reviews document[60] outlines the procedures for the CDA-AMC Reimbursement Review processes, including those used for oncology drugs, nononcologic drugs, and plasma protein and related products reviewed through the interim process.

References

1. O'Rourke B, Oortwijn W, Schuller T. The new definition of health technology assessment: A milestone in international collaboration. *Int J Technol Assess Health Care.* 2020;36(3):187-190. doi: 10.1017/s0266462320000215 PubMed
2. Haynes B. Can it work? Does it work? Is it worth it? The testing of healthcare interventions is evolving. *BMJ*. 1999;319(7211):652-3. doi: 10.1136/bmj.319.7211.652PubMed
3. EUnetHTA. EUnetHTA Joint Action 2 Work Package 8, HTA Core Model version 3.0. 2016. Accessed July 16, 2024. https://www.eunethta.eu/wp-content/uploads/2018/03/HTACoreModel3.0-1.pdf
4. Williamson PR, Altman DG, Bagley H, et al. The COMET Handbook: version 1.0. *Trials.* 2017;18(Suppl 3):280. doi: 10.1186/s13063-017-1978-4PubMed
5. Ciani O, Manyara AM, Davies P, et al. A framework for the definition and interpretation of the use of surrogate endpoints in interventional trials. *EClinicalMedicine.* 2023;65:102283. doi: 10.1016/j.eclinm.2023.102283PubMed
6. Ciani O, Buyse M, Drummond M, Rasi G, Saad ED, Taylor RS. Time to Review the Role of Surrogate End Points in Health Policy: State of the Art and the Way Forward. *Value Health.* 2017;20(3):487-495. doi: 10.1016/j.jval.2016.10.011PubMed
7. CADTH. Guidelines for the economic evaluation of health technologies: Canada. 4th ed. 2017. Accessed July 18, 2024. https://www.cadth.ca/guidelines-economic-evaluation-health-technologies-canada-4th-edition
8. National Institute for Health and Care Excellence. NICE health technology evaluations: the manual. (NICE process and methods PMG36). January 31, 2022. Updated October 31, 2023. Accessed July 16, 2024. https://www.nice.org.uk/process/pmg36
9. U.S. Food and Drug Administration. Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure. 2024. Accessed August 2, 2024. https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure
10. Kahan BC, Hindley J, Edwards M, Cro S, Morris TP. The estimands framework: a primer on the ICH E9(R1) addendum. *BMJ*. 2024;384:e076316. doi: 10.1136/bmj-2023-076316PubMed
11. Remiro-Azocar A. Some considerations on target estimands for health technology assessment. *Stat Med.* 2022;41(28):5592-5596. doi: 10.1002/sim.9566PubMed
12. CADTH. Reimbursement Review: sponsor summary of clinical evidence template, version 3. 2023. Accessed July 18, 2024. https://www.cadth.ca/sites/default/files/Drug_Review_Process/CADTH_Clinical_Evidence_Template.docx
13. Skoetz N, Goldkuhle M, van Dalen EC, et al. GRADE guidelines 27: how to calculate absolute effects for time-to-event outcomes in summary of findings tables and Evidence Profiles. *J Clin Epidemiol.* 2020;118:124-131. doi: 10.1016/j.jclinepi.2019.10.015PubMed
14. McKenzie JE, Brennan SE. Higgins JPT, Thomas J, Chandler J, et al, eds. Chapter 12: Synthesizing and presenting findings using other methods. In: Cochrane Handbook for Systematic Reviews of Interventions version 6.4. Cochrane; 2023. Updated October 2019. Accessed November 30, 2023. https://training.cochrane.org/handbook/current/chapter-12
15. Popay J, Roberts H, Sowden A, et al. Guidance on the conduct of narrative synthesis in systematic reviews. ESRC Methods Programme; 2006. Accessed November 30, 2023.
16. U.S. Food and Drug Administration. Integrated summary of effectiveness: guidance for industry. 2015. Accessed July 19, 2024. https://www.fda.gov/media/72335/download
17. Deeks JJ, Higgins JPT, Altman DG, (editors). Higgins JPT, Thomas J, Chandler J, et al, eds. Chapter 10: Analysing data and undertaking meta-analyses. In: Cochrane Handbook for Systematic Reviews of Interventions version 6.4. Cochrane; 2023. Updated August 2023. Accessed July 19, 2024. www.training.cochrane.org/handbook
18. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi: 10.1136/bmj.n71 PubMed
19. Balshem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol.* 2011;64(4):401-6. doi: 10.1016/j.jclinepi.2010.07.015PubMed

20. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. *Value Health.* 2011;14(4):417-28. doi: 10.1016/j.jval.2011.04.002PubMed

21. Macabeo B, Quenechdu A, Aballea S, Francois C, Boyer L, Laramee P. Methods for Indirect Treatment Comparison: Results from a Systematic Literature Review. *J Mark Access Health Policy.* 2024;12(2):58-80. doi: 10.3390/jmahp12020006 PubMed

22. Chan AW, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med.* 2013;158(3):200-7. doi: 10.7326/0003-4819-158-3-201302050-00583 PubMed

23. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332. doi: 10.1136/bmj.c332 PubMed

24. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c869. doi: 10.1136/bmj.c869 PubMed

25. Boutron I, Page MJ, Higgins JPT, Altman DG, Lundh A, Hróbjartsson A. Higgins JPT, Thomas J, Chandler J, et al, eds. Chapter 7: Considering bias and conflicts of interest among the included studies. In: Cochrane Handbook for Systematic Reviews of Interventions version 6.4. Cochrane; 2023. Updated August 2022. Accessed July 19, 2024. https://training.cochrane.org/handbook/current/chapter-07

26. Sterne JAC, Savovic J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898. doi: 10.1136/bmj.l4898 PubMed

27. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919. doi: 10.1136/bmj.i4919 PubMed

28. National Institute for Health and Care Excellence. Appendix C: Methodology checklist: randomised controlled trials. In: The guidelines manual: appendices B–I. 2012. Accessed July 11, 2024. https://www.nice.org.uk/process/pmg6/resources/the-guidelines-manual-appendices-bi-pdf-3304416006853

29. Committee for Medicinal Products for Human Use (CHMP). Reflection paper on establishing efficacy based on single-arm trials submitted as pivotal evidence in a marketing authorisation: considerations on evidence from single-arm trials. European Medicines Agency; 2023. Accessed July 16, 2024. https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-establishing-efficacy-based-single-arm-trials-submitted-pivotal-evidence-marketing_en.pdf

30. EUnetHTA. EUnetHTA 21 – Individual Practical Guideline Document: D4.4 – outcomes (endpoints), version 1. 2023. Accessed December 1, 2023. https://www.eunethta.eu/wp-content/uploads/2023/01/EUnetHTA-21-D4.4-practical-guideline-on-Endpoints-v1.0.pdf

31. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* 2010;19(4):539-49. doi: 10.1007/s11136-010-9606-8 PubMed

32. U.S. Food and Drug Administration. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims 2009. Accessed December 1, 2023. https://www.fda.gov/media/77832/download

33. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010;63(7):737-45. doi: 10.1016/j.jclinepi.2010.02.006 PubMed

34. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res.* 2011;11(2):171-84. doi: 10.1586/erp.11.9 PubMed

35. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61(2):102-9. doi: 10.1016/j.jclinepi.2007.03.012 PubMed

36. EUnetHTA. Guideline: endpoints used in relative effectiveness assessment: surrogate endpoints, adapted version. 2015. Accessed December 4, 2023. https://www.eunethta.eu/wp-content/uploads/2018/01/Endpoints-used-in-Relative-Effectiveness-Assessment-Surrogate-Endpoints_Amended-JA1-Guideline_Final-Nov-2015.pdf

37. EUnetHTA. Guideline: endpoints used in relative effectiveness assessment of pharmaceuticals: surrogate endpoints, final version. 2013. Accessed December 4, 2023. https://www.eunethta.eu/wp-content/uploads/2018/01/Surrogate-Endpoints.pdf
38. Tanniou J, van der Tweel I, Teerenstra S, Roes KC. Subgroup analyses in confirmatory clinical trials: time to be specific about their purposes. *BMC Med Res Methodol.* 2016;16(1):20. doi: 10.1186/s12874-016-0122-6 PubMed
39. Committee for Medicinal Products for Human Use (CHMP). Guideline on the investigation of subgroups in confirmatory clinical trials. European Medicines Agency; 2019. Accessed November 30, 2023. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-subgroups-confirmatory-clinical-trials_en.pdf
40. Schandelmaier S, Briel M, Varadhan R, et al. Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses. *CMAJ*. 2020;192(32):E901-E906. doi: 10.1503/cmaj.200077 PubMed
41. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328(7454):1490. doi: 10.1136/bmj.328.7454.1490 PubMed
42. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol.* 2011;64(4):383-94. doi: 10.1016/j.jclinepi.2010.04.026 PubMed
43. Dias S, Sutton AJ, Welton NJ, Ades AE. NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment. Decision Support Unit, ScHARR, University of Sheffield; 2011. Updated April 2012. Accessed July 19, 2024. https://www.sheffield.ac.uk/media/34179/download?attachment
44. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 1: Introduction to evidence synthesis for decision making. Decision Support Unit, ScHARR, University of Sheffield; 2011. Updated April 2012. Accessed July 19, 2024. https://www.sheffield.ac.uk/media/34175/download?attachment
45. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. Decision Support Unit, ScHARR, University of Sheffield; 2011. Updated September 2016. Accessed July 19, 2024. https://www.sheffield.ac.uk/media/34176/download?attachment
46. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials. Decision Support Unit, ScHARR, University of Sheffield; 2011. Updated April 2014. Accessed July 19, 2024. https://www.sheffield.ac.uk/media/34181/download?attachment
47. Phillippo D, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. NICE DSU technical support document 18: methods for population-adjusted indirect comparisons in submissions to NICE. 2016. Accessed July 19, 2024. https://www.sheffield.ac.uk/media/34216/download?attachment
48. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. *Value Health.* 2011;14(4):429-37. doi: 10.1016/j.jval.2011.01.011 PubMed
49. Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med.* 2015;162(11):777-784. doi: 10.7326/M14-2385 PubMed
50. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358:j4008. doi: 10.1136/bmj.j4008 PubMed
51. Whiting P, Savović J, Higgins JP, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol.* 2016;69:225-34. doi: 10.1016/j.jclinepi.2015.06.005 PubMed
52. CADTH, Health Canada. Guidance for reporting real-world evidence. 2023. Accessed July 18, 2024. https://www.cadth.ca/sites/default/files/RWE/MG0020/MG0020-RWE-Guidance-Report-Secured.pdf
53. Zisis K, Pavi E, Geitona M, Athanasakis K. Real-world data: A systematic literature review on the barriers, challenges, and opportunities associated with their inclusion in the health technology assessment process [non peer-reviewed preprint]. *medRxiv*. doi: 10.1101/2023.10.18.23297151

54. Fleurence RL, Kent S, Adamson B, et al. Assessing Real-World Data From Electronic Health Records for Health Technology Assessment: The SUITABILITY Checklist: A Good Practices Report of an ISPOR Task Force. *Value Health.* 2024;27(6):692-701. doi: 10.1016/j.jval.2024.01.019 PubMed

55. Schneeweiss S, Rassen JA, Brown JS, et al. Graphical Depiction of Longitudinal Study Designs in Health Care Databases. *Ann Intern Med.* 2019;170(6):398-406. doi: 10.7326/M18-3079 PubMed

56. Santesso N, Glenton C, Dahm P, et al. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. *J Clin Epidemiol.* 2020;119:126-135. doi: 10.1016/j.jclinepi.2019.10.014 PubMed

57. HTAGlossary.net. Accessed July 16, 2024. https://htaglossary.net/

58. Benkhalti M, Espinoza M, Cookson R, Welch V, Tugwell P, Dagenais P. Development of a checklist to guide equity considerations in health technology assessment. *Int J Technol Assess Health Care.* 2021;37:e17. doi: 10.1017/s0266462320002275 PubMed

59. Bond K, Stiffell R, Ollendorf DA. Principles for deliberative processes in health technology assessment. *Int J Technol Assess Health Care.* 2020;36(4):445-452. doi: 10.1017/S0266462320000550 PubMed

60. Canada's Drug Agency. Procedures for reimbursement reviews, version 22. 2024. Accessed July 18, 2024. https://www.cadth.ca/sites/default/files/Drug_Review_Process/CADTH%20Drug%20Reimbursement%20Review%20Procedures.pdf

61. U.S. Food and Drug Administration. E9(R1) statistical principles for clinical trials: addendum: estimands and sensitivity analysis in clinical trials: guidance for industry. 2021. Accessed July 30, 2024. https://www.fda.gov/media/148473/download

62. Westreich D, Edwards JK, Lesko CR, Cole SR, Stuart EA. Target Validity and the Hierarchy of Study Designs. *Am J Epidemiol.* 2019;188(2):438-443. doi: 10.1093/aje/kwy228 PubMed

63. Ades AE, Welton NJ, Dias S, Phillippo DM, Caldwell DM. Twenty years of network meta-analysis: Continuing controversies and recent developments. *Res Synth Methods.* 2024;15(5):702-727. doi: 10.1002/jrsm.1700 PubMed

64. Canada's Drug Agency. Expert Committee Deliberation at Canada's Drug Agency. 2025. Accessed March 3, 2025. https://www.cda-amc.ca/sites/default/files/MG%20Methods/expert_committee_deliberation.pdf

65. Canada's Drug Agency. Reimbursement Reviews. Updated October 29, 2020. Accessed July 30, 2024. https://www.cadth.ca/reimbursement-reviews

Appendix 1: Glossary

Please note that this appendix has not been copy-edited.

Note that the terms and definitions in this glossary may change over time as language and vocabulary continues to evolve. The terms and definitions included in this list may not be standardized, and their use may vary between individuals, groups, and regions.

The HTA Glossary may also be helpful.

Table 6: Glossary of Terms

Term	Definition
Appraisal of evidence, critical appraisal	"The process of assessing and interpreting scientific research results by systematically analysing their validity, clinical and statistical significance, and clinical relevance."[57]
Assessment	"A scientific process used to describe and analyse the properties of a health technology—its safety, efficacy, feasibility and indications for use, cost and cost-effectiveness, as well as social, economic and ethical consequences."[57]
Consistency	The agreement between the direct and indirect estimates of the treatment effects, that is, that the direct and indirect evidence sources estimate the exact same parameters. Consistency relies on the assumption of exchangeability: if this assumption holds then the direct and indirect estimates should be similar.[63]
Drug product	Products eligible for review, or under review, by CDA-AMC. Refer to Procedures for Reimbursement Reviews[60] for eligible products.
Effectiveness	The effect of a drug (or other technology) observed under routine conditions (in contrast to efficacy).
Efficacy	The effect of a drug (or other technology) observed under ideal conditions, such as a clinical trial (in contrast to effectiveness).
End point	"An indicator chosen for determining the effect of an intervention."[57]
Estimand	A precise description of the treatment effect that reflects the research question in a clinical trial. It includes the following attributes: treatment and comparator treatment, population, end point, how intercurrent events will be handled in the analysis, and the population summary for the end point.[61] A single study may have several estimands.
Head-to-head study or trial or evidence	A randomized controlled trial that includes a drug or other technology under review and another drug or technology as a comparator.
Homogeneity	Qualitative homogeneity (also referred to as exchangeability) exists if each trial estimates the same single treatment effect or different treatment effects distributed around a typical value; quantitative homogeneity is tested using statistical methods like the Q-test and the I^2 heterogeneity measure or comparing the fit of fixed and random effects models. Quantitative heterogeneity is what is measured by between-trials variance, while heterogeneity refers to variation within treatment comparisons.[63]
Index date	Generally, this refers to the start of the observational period in a retrospective study of administrative or other health care data. In a comparative effectiveness study of administrative health care data, it would generally refer to the start of exposure to a drug. The index date can vary and should be defined in each study.

Term	Definition
Intermediate outcome	A clinical end point, such as a measure of a function or of a symptom (i.e., disease-free survival, symptom frequency, functional capacity), but not the ultimate end point of the disease, such as survival or the rate of irreversible morbid events. Improvement in an intermediate outcome due to treatment is well perceived and can be of value to patients even if it does not lead to improvement of morbidity or mortality.
Pivotal trial	A study designed to support the efficacy and safety of a drug for a regulatory submission.
Reimbursement review	Reimbursement Reviews performed by CDA-AMC are comprehensive assessments of the clinical effectiveness and cost-effectiveness, as well as patient and clinician perspectives, of a drug or drug class. The assessments inform nonbinding recommendations that help guide the reimbursement decisions of Canada's federal, provincial, and territorial governments, with the exception of Quebec.[65]
Similarity	The degree to which the trials included in the indirect comparison have comparable populations, interventions, and outcomes per the PICO[T][S] framework; implied by exchangeability.[63]
Surrogate outcome	A biomarker or intermediate outcome used to substitute for a patient-relevant final (or target) outcome that reliably predicts benefit or harm based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence. This definition includes both biomarkers (e.g., blood pressure, tumour response), and intermediate outcomes (such as PFS), which may have potential direct relevance to patients.[6]
Target (final) outcome	Target outcomes of interest in HTA are those that estimate the clinical benefit and thereby help estimate the clinical value of the drug product. The assessment of health benefits considers clinically meaningful end points such as mortality, morbidity, and patient-reported experiences and feelings, symptoms, health behaviours, function, and health-related quality of life.[3]

CDA-AMC = Canada's Drug Agency; HTA = health technology assessment; PFS = progression-free survival.

Canada's Drug Agency
L'Agence des médicaments du Canada
Drugs. Health Technologies and Systems. Médicaments, technologies de la santé et systèmes.

Canada's Drug Agency (CDA-AMC) is a pan-Canadian health organization. Created and funded by Canada's federal, provincial, and territorial governments, we're responsible for driving better coordination, alignment, and public value within Canada's drug and health technology landscape. We provide Canada's health system leaders with independent evidence and advice so they can make informed drug, health technology, and health system decisions, and we collaborate with national and international partners to enhance our collective impact.

Questions or requests for information about this report can be directed to Requests@CDA-AMC.ca.