Comparison of Value Framework Assessments for Multiple Myeloma

Final White Paper

HEALTHCARE AND HUMAN SERVICES POLICY, RESEARCH, AND CONSULTING—WITH REAL-WORLD PERSPECTIVE.

Prepared for: **National Pharmaceutical Council**

Submitted by: **The Lewin Group, Inc.**

November 2016

Table of Contents

This report was prepared by Christel Villarivera and Clifford Goodman of The Lewin Group.

I. Introduction

Advances in cancer therapies have led to improved clinical outcomes, including longer survival and improved quality of life in some cancers, but have also been accompanied by increased spending in oncology, including on cancer drugs and biologics. As more expensive targeted cancer treatments become standards of care, the costs of cancer care are expected to continue to rise. It is projected that there will be 18.1 million cancer survivors in 2020 in the U.S. (30% more than in 2010), and that the costs of cancer care will reach $157.7 billion.[1] In recent years, drugs that have been approved by the U.S. Food and Drug Administration (FDA) for cancer indications are routinely priced at $10,000 or more per month and greater than $100,000 per year of therapy.[2] For cancer patients, treatment costs are significant, with typical out-of-pocket expenses of $20,000 to $30,000 a year, approaching half of the median annual household income in the U.S. This cost burden contributes to findings that an estimated 10% to 20% of patients compromise their treatment plans or decide not to receive treatment.[3,4]

Value assessment frameworks have emerged as tools to respond to demand expressed by patients, clinicians, payers, and other stakeholders for greater health care decision support, particularly to define and measure the relative value of treatment options. In addition to matters of weighing benefits and harms of treatment options and higher prices, some of the broader contextual factors that have influenced the development of these tools include: the continued shift in the basis of health care payment from volume to value and accompanying alternative payment models designed to incentivize delivery of value, the growing capacity to generate real-world evidence of value, advances in personalized medicine, and an increased focus on patient- and consumer-centered health care.

Three of the value assessment frameworks being developed in the U.S. specifically focus on cancer treatments, primarily cancer drugs and drug regimens:

- The American Society of Clinical Oncology (ASCO) Conceptual Framework to Assess the Value of Cancer Treatment Options[5]
- Memorial Sloan Kettering Cancer Center‘s (MSKCC) DrugAbacus[6]
- The National Comprehensive Cancer Network (NCCN) Evidence Blocks™[7]

[1] AB, Yabroff KR, Shao Y, Feuer EJ, Brown ML. Projections of the cost of cancer care in the United States: 2010-2020. J Natl Cancer Inst. 2011 Jan 19;103(2):117-28.
[2] Light DW, Kantarjian H. Market spiral pricing of cancer drugs. Cancer. 2013 Nov 15;119(22):3900-2.
[3] Kantarjian H, Steensma D, Rius Sanjuan J, Elshaug A, Light D. High cancer drug prices in the United States: reasons and proposed solutions. J Oncol Pract. 2014 Jul;10(4):e208-11.
[4] Howard DH, Bach PB, Berndt ER, Conti RM. Pricing in the market for anticancer drugs. J Econ Perspect. 2015;29(1):139-62.
[5] Schnipper LE, Davidson NE, et al. American Society of Clinical Oncology Statement: A Conceptual Framework to Assess the Value of Cancer Treatment Options. J Clin Oncol. 2015 Aug;33(23):2563-77.
[6] Memorial Sloan Kettering Cancer Center. DrugAbacus FAQs. DrugAbacus website. http://www.drugabacus.org/faqs. Accessed March 4, 2016.
[7] National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology (NCCN Guidelines®) with NCCN Evidence Blocks™. NCCN website. http://www.nccn.org/evidenceblocks/. Accessed July 11, 2016.

Further, the value framework developed by the Institute for Clinical and Economic Review (ICER) has been used to assess the value of certain cancer treatments, though it is not limited to cancer treatments.[8]

The intended users of these value frameworks differ. The ASCO and NCCN value assessment tools focus on assisting shared decision-making by cancer clinicians and their patients. The DrugAbacus tool and ICER value assessment framework assess value at a broader level and are more relevant to payers and policymakers. Even so, these tools are available to patients and clinicians.

There are certain common themes or domains across these tools, including those of clinical benefit, toxicity, and some economic component. However, comparisons of these frameworks highlight important differences and analytic challenges, including the different methodologies, inputs, outputs, and assumptions of each that yield important differences in results, even between frameworks with the same main intended purpose and audiences. Users encounter incongruent formats and findings across these frameworks and may have difficulty in determining whether or how to apply them to their particular needs. At the same time, there is potential utility in multiple frameworks that address the same or similar topics from different perspectives.

II. Purpose

The Lewin Group (Lewin) was contracted by the National Pharmaceutical Council (NPC) to conduct an independent analysis to compare how different value frameworks assessed treatments for multiple myeloma. This topic was selected because four value frameworks have assessed or otherwise addressed treatments for myeloma, and therefore it can serve as an illustrative example of how these frameworks compare.

The analysis addresses the frameworks' purposes, audiences, methodologies, and results. Further, it considers the underlying reasons for observed differences among the frameworks and the implications of these differences for users. This analysis also identifies opportunities to address these differences and to encourage further development, including where certain standards across frameworks might help to advance the field.

This analysis recognizes that these four frameworks are in different stages of development. Indeed, although some are more developed than others, each of their developers considers their framework to be a work in progress. All of these frameworks have evolved in response to feedback from various stakeholders.

[8] Institute for Clinical and Economic Review. Value assessment project, a framework to guide payer assessment of the value of medical services.
ICER website. http://www.icer-review.org/impact-and-outcomes/value-assessment-project/. Published September 2015. Accessed March 4, 2016.

III. Methodology for this Analysis

Lewin began by reviewing and gathering publicly available information about each of the frameworks from their developers. We also examined information from published and other publicly available sources, including comparative reviews that have been conducted and any relevant grey literature, including NPC's *Current Landscape: Value Assessment Frameworks (2016)*.

Following a prior analysis Lewin conducted for NPC, *Comparison of Value Assessment Frameworks Using the National Pharmaceutical Council's Guiding Practices for Patient-Centered Value Assessment (2016)*, which examined value assessment frameworks using NPC's *Guiding Practices for Patient-Centered Value Assessment*, we examined each framework's assessment of multiple myeloma treatments and captured information from each. The main types of information captured were patient population(s)/indications of interest, treatments/regimens of interest, outcomes/parameters assessed (health, economic, other), and overall findings. We also reviewed other sources that had raised issues about some of these multiple myeloma assessments and responses that the developers provided in the public domain to address such issues.

We interviewed representatives from each of the four organizations that developed the value assessment frameworks via teleconference. These calls and multiple follow-up communications enabled further understanding about the purpose or scope of their frameworks and other information that may not have been available in the public domain. Lewin also interviewed relevant stakeholders and experts, including in the areas of oncology and multiple myeloma, to gain their perspectives on value assessment, the application of the frameworks to multiple myeloma treatments, and additional context to the analysis.

IV. Findings

A. Framework Assessments of Multiple Myeloma

The following section describes each value assessment framework or tool and how they were applied to assess drug regimens for multiple myeloma. Table 1 presents the drug regimens examined by each framework for advanced multiple myeloma not previously treated and for relapsed or refractory multiple myeloma.

Table 1. Regimens Examined by Each Framework

Regimen	ASCO	DrugAbacus	NCCN*	ICER
Advanced multiple myeloma not previously treated				
Bortezomib + melphalan, and prednisone (BOR+MEL+PRED)	X		X	
Bortezomib + lenalidomide + dexamethasone (BOR+LEN+DEX)			X	
Lenalidomide + low-dose dexamethasone (LEN+loDEX)			X	
Melphalan + prednisone + lenalidomide (MEL+PRED+LEN)			X	
Melphalan + prednisone + thalidomide (MEL+PRED+THAL)			X	

Regimen	ASCO	DrugAbacus	NCCN*	ICER
Relapsed or refractory multiple myeloma				
Bortezomib (BOR)		X	X	
Carfilzomib + lenalidomide + dexamethasone (CFZ+LEN+DEX)			X	X
Elotuzumab + lenalidomide + dexamethasone (ELO+LEN+DEX)			X	X
Ixazomib + lenalidomide + dexamethasone (IX+LEN+DEX)			X	X
Daratumumab monotherapy (DARA)			X	X
Panobinostat + bortezomib + dexamethasone (PAN+BOR+DEX)		X	X	X
Pomalidomide + low-dose dexamethasone (POM+LoDEX)		X	X	X
Lenalidomide + dexamethasone (LEN + DEX)			X	
Bortezomib + liposomal doxorubicin (BOR + liposomal DEX)			X	

* With the exception of DARA, listed regimens are NCCN preferred regimens with Category I evidence and consensus. DARA is an NCCN preferred regimen but is not classified as Category 1.

1. ASCO

ASCO's value assessment framework, which examines the relative value of cancer therapies that have been compared in clinical trials, is still being developed. To date, ASCO has released two versions. The initial version was published in June 2015, and version 2.0 was published in May 2016, reflecting input and feedback that ASCO received following release of the initial version. ASCO plans to develop a user-friendly software tool based on the framework for clinicians to use with their patients to help inform their treatment decisions.

ASCO's framework yields a composite net health benefit (NHB) score for a drug regimen relative to the comparator from a clinical trial. The initial version of the framework is described below as the multiple myeloma assessment conducted by ASCO utilized the initial framework. The NHB is based on:

- points for clinical benefits (e.g., improvement in overall survival [OS], progression-free survival [PFS], or response rate)
- positive or negative points for toxicity (grade 3-5)
- bonus points for symptom palliation and treatment-free survival.

The NHB score is intended to support a provider and patient in assessing the additional benefit of the drug regimen of interest compared with the standard of care. The maximum possible score for clinical benefit is 80 points; the maximum possible score for toxicity is 20 points (where a higher score indicates a better tolerated drug, i.e., with less toxicity); and the maximum number of bonus points that can be awarded is 30. As such, the maximum NHB score based on these scores is 130 points. In this framework, the higher the NHB is, the greater the additional clinical benefit and/or lower the toxicity associated with the drug regimen of interest is relative to the comparator. An NHB of zero reflects equivalence of the drug regimen of interest and the comparator (not that the drug regimen of interest is ineffective).

Costs are provided separately from the NHB score. For this, ASCO uses drug acquisition costs based on average sales price as of October 2014 for intravenous therapies and on information

from UnitedHealthcare for oral drugs.[9] The initial framework sought to address patient copays but indicated that individual patient copay amounts were to be determined for each patient at the point of service, given that patient copays are highly variable depending on insurance plans and other factors.

ASCO described certain limitations of the NHB score when the initial framework was published, including that the NHB for a given regimen is only meaningful within the context of the trial, relative to the comparator. As such, NHB scores cannot and should not be compared to each other. Another limitation is that study populations must meet certain requirements to be eligible for a trial and are unlikely to represent the general cancer population.

The most notable revisions reflected in version 2.0 of the framework include: the calculation of treatment efficacy now uses hazard ratios when available rather than absolute improvements in survival; consideration of more reported side effects, not just the most severe, high-grade toxicities; and the addition of bonus points for improvement in quality of life (QoL) and significant improvement in survival at the tail end of the curve.[10]

ASCO noted feedback that the NHB score was viewed as being arbitrary, not intuitive, and lacking the meaning that an absolute value for clinical benefit or toxicity would have. Although ASCO concurred that NHB is an artificial construct, it went on to state that the NHB is derived from the key efficacy elements of OS, PFS, RR, symptom palliation, time off treatment, and QoL, as well as the comparative toxicity of the regimen and that these are the elements that oncologists consider when making treatment recommendations and that patients try to understand as they consider their treatment options. ASCO also re-emphasized that the NHB score serves as an indicator of the relative clinical impact and toxicity of the test regimen of interest as compared with a comparator regimen. ASCO indicates that it is important for providers to make clear the absolute magnitude of benefit that their patient might expect from the therapy under consideration, to minimize the chance for misinterpretation.

To assess the utility of the initial version of their framework and to inform refinements to the framework, ASCO applied it to four clinical case scenarios. Early versions of the framework as well as these clinical scenarios were shared with a range of stakeholders for input before ASCO published the initial framework. One of three clinical case scenarios that ASCO presented with the first version of its framework was the primary treatment of advanced multiple myeloma.[11,12]

ASCO's assessment of treatments for multiple myeloma using the initial framework relied on data from a single head-to-head randomized controlled trial (RCT) that was referenced in ASCO's publication, the VISTA (Velcade as Initial Standard Therapy in Multiple Myeloma: Assessment With Melphalan and Prednisone) trial. The published study was further examined to

[9] Schnipper et al. 2015.

[10] Tail of the curve bonus points are awarded if: (1) the test regimen results in at least a 50% relative improvement in the proportion of patients who are alive with the test regimen at the time point on the survival curve that is at twice the median OS or PFS point for the control regimen, and (2) at least 20% of patients receiving the control regimen are alive at this time.

[11] San Miguel JF, Schlag R, Khuageva NK, et al. Bortezomib plus melphalan and prednisone for initial treatment of multiple myeloma. N Engl J Med 359:906-917, 2008.

[12] San Miguel JF, Schlag R, Khuageva NK, et al. Persistent overall survival benefit and no increased risk of in patients with previously untreated multiple myeloma. J Clin Oncol 31:448-455, 2013.

better understand the methodology that ASCO applied. The target population was patients with advanced multiple myeloma who were not previously treated (n=682). The intervention of interest was the drug regimen of bortezomib, melphalan, and prednisone (BOR+MEL+PRED) compared to MEL+PRED (the control). At the time of the trial, MEL+PRED was the standard treatment for patients with multiple myeloma who were not candidates for high-dose therapy. In order to calculate the NHB, ASCO's framework took into account the differences in median OS, the frequency of high-grade toxicities, and median treatment-free survival.

- The median OS was 56.4 months for BOR+MEL+PRED versus 43.1 months for MEL+PRED, indicating a 31% improvement in median OS as a result of the intervention. Per ASCO's framework for advanced disease, this percentage change in median OS fell within the range of an OS score of 2 (scores range from 1-5), which ASCO's framework instructed should be multiplied by 16 to reach the total assessment points for OS, which equaled 32 points.
- A total of 42 grade 3-5 toxicities were observed in the BOR+MEL+PRED group versus 34 grade 3-5 toxicities in the MEL+PRED group, which represents an increase in toxicity of 24% in the intervention group. As per ASCO's framework, if there is less than a 49% increase in toxicity (or if there is a less than 49% decrease in toxicity), the toxicity is considered the same for the intervention and comparator, and the toxicity score is zero.
- The median treatment-free survival was 30.7 months for BOR+MEL+PRED versus 20.5 months for MEL+PRED, indicating a 50% improvement in median treatment-free survival as a result of the intervention. As per ASCO's framework, this percentage difference is awarded 15 bonus points.

Based on the clinical benefit score, the toxicity score, and the bonus points, the NHB score for BOR+MEL+PRED when compared to the standard treatment was 47 points out of a maximum of 130 points. ASCO reported separately that the average cost of the intervention was $7,042.70 per month, and the average cost of the comparator was $279.45 per month, a difference of more than $6,700 per month. ASCO indicated that patient copay costs were to be evaluated individually at the point of care.

When ASCO published version 2.0 of its framework, in addition to two new clinical scenarios, ASCO applied the revised framework to two of the four clinical scenarios to which it had applied the initial version in order to demonstrate the differences in NHB calculations as a result of the revisions that were made. However, treatment for advanced multiple myeloma was not one of the two clinical scenarios ASCO revisited.[13]

2. DrugAbacus

MSKCC's DrugAbacus is an interactive online tool that contains data on a large though not systematically selected set of 52 drugs approved between 2001 and 2015 by the FDA for the treatment of cancer. DrugAbacus allows users to weight their preferences across a set of value domains. The initial tool accounted for the following six value domains: efficacy, tolerability, novelty, research and development costs, rarity, and population burden. In July 2016, stakeholder

[13] Schnipper et al. 2015.

feedback led to the inclusion of two additional domains, unmet need and prognosis, which were then incorporated into the updated tool and can also be weighted based on the user's preferences. The tool yields a value-based monthly price, called a DrugAbacus Price, calculated using a formula that consists of the set of domain parameters weighted based on the user's preferences. The resulting DrugAbacus price is juxtaposed versus the actual market monthly price (i.e., cost to Medicare) for comparison.

The DrugAbacus equation is as follows:

$$\text{Price} = (\beta\_\text{eff}\ [\![\cdot X]\!]\_\text{eff})(1-(\beta\_\text{tol}\ [\![\cdot X]\!]\_\text{tol}))(\beta\_\text{nov}\ [\![\cdot X]\!]\_\text{nov})(\beta\_(R\&D)\ [\![\cdot X]\!]\_(R\&D)(\beta\_\text{rar}\ [\![\cdot X]\!]\_\text{rar})(\beta\_\text{bur}\ [\![\cdot X]\!]\_\text{bur})\ (\beta\_\text{unm}\ [\![\cdot X]\!]\_\text{unm})(\beta\_\text{prg}\ [\![\cdot X]\!]\_\text{prg}))$$

The DrugAbacus website indicates that each X is a measurement for one of the eight domains (efficacy, tolerability, novelty, research & development costs, rarity, population burden, unmet need, and prognosis). These values are extracted by the DrugAbacus team based on the clinical trial data and/or market profile for the first indication that was FDA-approved for each drug included in the tool. However, DrugAbacus does not provide the actual values it uses for each drug and does not provide references for the clinical trials that were used. While the DrugAbacus website notes that the drugs included in the tool were those examined by Howard et al., it does not provide a citation for this study.

The website also indicates that each β is a weight defining the importance of that domain with respect to the drug's price as selected by the user within a predetermined range of possible values. These values are obtained from user-adjusted sliders within the online tool. Table 2 presents the range of values for each domain's slider.

Table 2. Drug Abacus Domains

Domain	Definition Provided to User	Range of Value
Dollars Per Year	Price the Abacus should use for a year of life	$12,000 - $300,000
Toxicity Discount	Maximum discount the Abacus should apply to drugs with severe side effects	0% - 30%
Novelty Multiplier	Maximum premium the Abacus should apply to drugs with novel mechanisms of action	1.0 - 3.0
Rarity Multiplier	Maximum premium the Abacus should apply to drugs that treat rare illnesses	1.0 - 3.0
Population Burden of Disease	Maximum premium the Abacus should apply to drugs that address large population health burdens	1.0 - 3.0
Cost of Development	Maximum premium the Abacus should apply to drugs that are expensive to develop	1.0 - 3.0
Prognosis	Maximum premium the Abacus should apply to drugs that treat aggressive illnesses	1.0 - 3.0
Unmet Need	Maximum premium the Abacus should apply to drugs that treat illnesses for which there are few or no other treatments available	1.0 - 3.0

According to its developers, although DrugAbacus is often grouped with or compared to value assessment frameworks, it is primarily a research tool meant to explore and test different concepts that could affect a drug's value. The DrugAbacus terms and conditions, which users must indicate they have read and agreed to before accessing the tool, include the following:

> "DrugAbacus is a research tool created for general information only. The information is not to be used as a substitute for medical advice, diagnosis, or treatment of any health condition or problem, nor is it a substitute for primary pharmacoeconomic or clinical efficacy assessment. Users of DrugAbacus should not rely on information provided by it for decisions about pricing, insurance coverage policy, or forecasting, or about the benefits, harms, or other attributes of individual listed drugs."[14]

A search of the DrugAbacus tool identified three drugs used to treat multiple myeloma. DrugAbacus prices for the following drugs are included: Farydak (panobinostat, PAN), Pomalyst (pomalidomide, POM), and Velcade (bortezomib, BOR). However, the tool provides no additional information for these drugs beyond monthly cost, limiting the transparency of the tool. A separate search of these drugs on the FDA's website was conducted for this analysis in order to determine the indications for which they were first approved by the FDA. The search found that bortezomib was first approved for the treatment of patients with multiple myeloma who had received at least one prior therapy (i.e., as a second-line treatment). Panobinostat and pomalidomide were first approved for the treatment of patients with multiple myeloma whose cancer has progressed after treatment with at least two prior standard therapies (i.e., as third-line treatments).

While DrugAbacus is transparent about the general equation used to determine DrugAbacus prices, the actual data and data sources used to calculate the Abacus price for each drug in the tool are not made readily accessible to users and are not referenced, making it difficult to reproduce the Abacus price. However, through detailed exploration of the source code of the website, and the ability to decipher this code, it appears to be possible to retrieve data used to calculate each drug's Abacus price. Still, it is unlikely that DrugAbacus expected users would examine the source code for its website, and that users would think to examine the source code in order to access these data.

In examples where DrugAbacus is described, midrange estimates for a year of life and toxicity discount are often used, with other domains set to a value of 1.0, to calculate a midrange DrugAbacus price. Table 3 shows the values for PAN, POM, and BOR when setting the value of a year of life at $132,000, with a toxicity discount that subtracts 15% off that level, and with all remaining sliders set to 1.0 (i.e., no effect); of the three drugs, the estimated DrugAbacus price for BOR is the only one that exceeds the actual monthly price using these assumptions. Table 3 also presents the DrugAbacus prices when setting all sliders in the tool to their minimum and maximum values for these drugs. Lastly, the table presents the actual monthly prices (as paid by Medicare), which are provided in the DrugAbacus tool for comparison.

Table 3. Abacus Prices and Estimated Actual Monthly Prices

Drug	Midrange Estimate	Minimum	Maximum	Actual Monthly Price
PAN	$7,817	$661	$169,006	$10,625
POM	$11,000	$1,000	$242,678	$14,165
BOR	$9,394	$841	$728,361	$4,474

[14] http://www.drugabacus.org/terms-conditions/

DrugAbacus is the only tool in this analysis that currently allows users to weight their preferences across a set of domains. While some of these domains are also common to certain ones considered by other frameworks in this analysis (e.g., efficacy, tolerability), other DrugAbacus domains are not considered by others, or not to the same extent as DrugAbacus (e.g., novelty, research and development costs, rarity, population burden, unmet need, prognosis).

Although there does not appear to be a formal process in place for it, DrugAbacus allows for feedback and input, and has responded with some changes. The recent update, which added the domains of unmet need and prognosis, also included the new feature of indication-based pricing, but only for four drugs: Afinitor, Avastin, Nexavar, and Tarceva. There are no plans at this time to update DrugAbacus. Its developers stated that there are no plans to add additional drugs to the tool unless there is a need to test new concepts.

3. NCCN

NCCN presents "Evidence Blocks" to accompany its oncology clinical practice guidelines. These five-by-five depictions represent five domains: efficacy (E), safety (S), quality of evidence (Q), consistency of evidence (C), and affordability (A). NCCN guideline panel members score each of the first four domains using a standardized scale of 1 (least favorable) through 5 (most favorable). For these domains, panel members rely on their knowledge of the published data, evidence cited in the NCCN guidelines, and their clinical experience with the treatments. Panel members score affordability based on their knowledge of the overall cost of the regimen, ranging from 1 (very expensive) to 5 (very inexpensive). Affordability is intended to account for costs including and associated with oncology therapies; this includes drug cost, supportive care, administration costs, and monitoring and management of toxicity. An average of the panel members' scores for each domain is calculated and rounded to the closest whole number, and these average scores are used to build the Evidence Blocks.

NCCN generates the initial evidence blocks using an emailed survey once the panel has met and made decisions about what to include in its guideline. Thereafter, panel members review the evidence blocks as part of their annual guideline review and discuss any new issues that may pertain to the blocks. If there are significant questions, the panel votes again, either in person or via an emailed survey depending on the time available and number of blocks with questions. As a component of NCCN's guidelines, the Evidence Blocks are intended to supply the provider and individual patient with information about value to support shared decision-making about care options.

At the time of our analysis the most recent version of the NCCN guidelines for multiple myeloma was released in April 2016. These guidelines, with their corresponding Evidence Blocks, are more comprehensive than the other frameworks being analyzed in that they address diagnosis and treatment of active multiple myeloma as well as solitary plasmacytoma (when there is only a single mass of myeloma cells) and smoldering multiple myeloma (asymptomatic multiple myeloma).[15]

[15] NCCN Clinical Practice Guidelines in Oncology. Multiple Myeloma NCCN Evidence Blocks Version 3. 2016. Ft. Washington, PA: National Comprehensive Cancer Network, April 2016.

For active multiple myeloma, the guidelines address preferred and other primary treatment regimens for transplant and non-transplant candidates, preferred and other primary treatment regimens for progressive or relapsed multiple myeloma, and adjunctive treatments. Evidence Blocks are provided for each of these regimens. With the Evidence Blocks for preferred regimens, NCCN indicates if a regimen received a "Category 1" rating based on NCCN's categories of evidence and consensus, which indicates that there was uniform NCCN consensus that the regimen is appropriate based on high-level evidence.

For the primary treatment of active multiple myeloma patients who are not transplant candidates, Evidence Blocks are provided for 13 different preferred regimens. Of these, the following five were classified as having Category 1 evidence and consensus and had the following Evidence Block scores:

Table 4. Evidence Block Scores for Primary Treatments for Multiple Myeloma

Preferred Regimen (Category 1)	E	S	Q	C	A
Bortezomib/lenalidomide/dexamethasone	5	3	4	4	1
Lenalidomide/low-dose dexamethasone	4	4	4	4	2
Melphalan/prednisone/bortezomib	4	3	4	4	3
Melphalan/prednisone/lenalidomide	4	3	4	4	2
Melphalan/prednisone/thalidomide	4	3	4	4	3

For relapsed/refractory multiple myeloma, Evidence Blocks are provided for a total of 23 different preferred regimens, the following 8 of which NCCN classified as having Category 1 evidence and consensus:

Table 5. Evidence Block Scores for Treatments for Relapsed/Refractory Multiple Myeloma

Preferred Regimen (Category 1)	E	S	Q	C	A
Bortezomib	3	4	4	4	2
Bortezomib/liposomal doxorubicin	4	3	4	4	2
Carfilzomib/lenalidomide/dexamethasone	5	3	4	4	1
Elotuzumab/lenalidomide/dexamethasone	3	3	4	4	1
Ixazomib/lenalidomide/dexamethasone	4	3	4	4	1
Lenalidomide/dexamethasone	4	4	4	4	2
Panobinostat/bortezomib/dexamethasone	3	2	4	4	2
Pomalidomide/dexamethasone	4	4	4	4	2

NCCN Evidence Blocks differ from the other frameworks in this analysis in that they currently supplement clinical practice guidelines that include a review of the available evidence. The Evidence Blocks are intended to inform the shared decision-making process of health care providers and their patients. While it does not provide actual weights for patient preferences, NCCN indicated that the Evidence Blocks can be used as a starting point for discussing a patient's preferences.

It is not fully apparent how the scores should be interpreted. NCCN provides some definitions for each 1-5 scale used for the five domains of the Evidence Blocks. However, while a 1-5 scale may be intuitive for most users, the basis of the scaling, e.g., whether it is simply ordinal or proportionate in some way, is not apparent. Similarly, for the domain of affordability, NCCN does not define or quantify what constitutes the 1 (very expensive) versus 5 (very inexpensive)

scores. NCCN considers feedback and requests for updates to their guidelines at any point in time and has a formal process for responding to feedback and requests. At present, it does not appear that NCCN has plans to update the methodology of the Evidence Blocks. NCCN has incorporated Evidence Blocks in at least 20 of its guidelines to date.

4. ICER

ICER's value assessment framework has evolved since its inception. The current version evaluates care value over the long term and potential health system budget impact over the short term. ICER also calculates what it calls "value-based price benchmarks," which account for long-term cost-effectiveness and the potential short-term budget impact of the treatment being assessed. ICER uses its value assessment framework to determine these benchmarks.

Care value accounts for comparative clinical effectiveness, incremental cost per outcomes achieved, other benefits or disadvantages, and contextual considerations. During a public meeting, members of one of ICER's three advisory panels vote on the care value results as high, intermediate, or low.

Regarding comparative clinical effectiveness, ICER considers both the magnitude of the comparative net health benefit and the level of certainty of the evidence supporting the net health benefit. For incremental cost-effectiveness, ICER emphasizes a long-term perspective on outcomes and costs. It does not consider patient out-of-pocket costs, but rather considers drug costs from a health care system or other payer perspective, including any related cost savings (e.g., prevention of hospitalization, reduced number of doctor visits, and public health benefits) over the long term. ICER applies a cost-effectiveness benchmark range (incremental cost-effectiveness ratio) of $100K-$150K per quality-adjusted life year (QALY) gained as an upper bound for what it considers to be generally reasonable value in the U.S. health care system.

ICER approaches patient groups and other stakeholders to gather information on other benefits or disadvantages to individual patients, their caregivers, the health care delivery system, and the public that may not have been captured or considered in the evidence on comparative clinical effectiveness (e.g., methods of administration that improve or diminish patient acceptability and adherence). ICER presents this information as part of its reports, and it is addressed during the public meetings of the advisory panels.

ICER also consults patient groups for information on contextual considerations, which include ethical, legal, or other issues that may influence the relative priority of illnesses and interventions. ICER includes this information in its reports, and it is discussed by its advisory panels during the public meetings. ICER has indicated that, in practice, these considerations often focus on conditions of very high severity for which acceptable treatments do not exist.

In addition to consideration of long-term care value, ICER's value assessment framework considers potential short-term budget impact. The potential budget impact accounts for the estimated net change (i.e., difference between the new treatment and the current/standard treatment) in total health care costs (including drug costs and potential impacts on hospital visits, doctor visits, other tests, etc.) over an initial five-year timeframe. In assessing the potential budget impact of a new treatment, ICER assigns one of four potential uptake patterns based on condition/market considerations (e.g., what treatments a new treatment may be replacing). The

uptake rates range from what ICER considers a low of 10% over five years to a very high uptake rate of 75% over five years, all in an unmanaged context. ICER does not intend these uptake rates to be estimates or projections of anticipated actual uptake rates.

ICER then examines whether the potential budget impact exceeds ICER's "alarm bell" threshold. ICER defines this threshold as the point at which the net cost increase for a new intervention would contribute to growth in overall health care spending annualized over five years that is greater than the anticipated growth in national GDP +1%. For 2015-2016, ICER calculated this threshold to be $904 million for an individual new drug. ICER states that this threshold is intended to serve as an "alarm bell" for a health care system considering whether utilization management, lower prices, reallocation of resources, or other consideration may be appropriate.

ICER uses the framework to estimate value-based price benchmarks under various scenarios. ICER first considers at what price a drug would achieve long-term cost effectiveness at the upper range of $100K-$150K per QALY. This range allows for the effects of other factors (i.e., other benefits and contextual considerations). Next, given an assumed uptake rate assigned to the drug over a five-year period, ICER determines the price at which the short-term potential budget impact would cross the alarm bell threshold of $904 million. ICER then presents each of these prices and the list wholesale acquisition cost (WAC) price of the drug for comparison. In some cases, the benchmark prices exceed the list price, and in other cases, the benchmark prices are less than the list price. ICER states that the benchmarks are not intended to be viewed as a right or fair price of a drug, but a way to call attention of decision-makers to matters of affordability of a new drug.

ICER released a draft report on treatment options for relapsed or refractory multiple myeloma in May 2016 using the initial version of the framework.[16] As part of its process, ICER provided for a two-week public comment period on the draft. ICER received public comments from certain multiple myeloma-focused organizations, manufacturers, and others. The main high-level concerns across the public comments included the following:

- Assessment of the effectiveness and sequencing of regimens did not adequately account for the heterogeneous subtypes of multiple myeloma.
- Comparisons of regimens were insufficiently reflective of clinical practice.
- Consideration of patient preferences and experience was inadequate.
- Analyses of clinical effectiveness and cost effectiveness were flawed.
- Assessment of multiple myeloma therapies was prematurely conducted when only limited evidence was available.
- The public comment period was too short for some organizations to respond adequately.

In preparing its final report, ICER responded to comments and related feedback on its draft report. Its full written response was made available on the ICER website. ICER issued the final

[16] Midwest CEPAC. Treatment Options for Relapsed or Refractory Multiple Myeloma: Effectiveness, Value, and Value-Based Price Benchmarks Evidence Report. Boston, MA: ICER, May 2016.

report and meeting summary in July 2016 following a public comment period and a public meeting of its independent advisory group.[17]

In its final report, ICER defined the population of interest as adults with multiple myeloma whose disease has not responded to at least one previous line of treatment (i.e., refractory) or has relapsed following such treatment, are not currently on maintenance treatment, and are not being considered for hematopoietic stem cell transplant.

The report included findings from ICER's clinical comparative effectiveness review.

1. The drug regimens of interest were:
    - Carfilzomib + lenalidomide + dexamethasone (CFZ+LEN+DEX)
    - Daratumumab monotherapy (DARA)
    - Elotuzumab + lenalidomide + dexamethasone (ELO+LEN+DEX)
    - Ixazomib + lenalidomide + dexamethasone (IX+LEN+DEX)
    - Panobinostat + bortezomib + dexamethasone (PAN+BOR+DEX)
    - Pomalidomide + low-dose dexamethasone (POM+LoDEX)
2. The comparator regimens were:
    - Lenalidomide + dexamethasone (LEN+DEX)
    - Bortezomib + dexamethasone (BOR+DEX)
3. The outcomes of interest included:
    - OS
    - Disease progression-related measures (PFS, time to progression)
    - Biochemical response (overall response rate)
    - Health-related quality of life (HRQoL)
    - Treatment-related adverse events:
        - Rates of grade 3 or 4 key adverse events
        - Rates of serious adverse events
        - Discontinuation due to adverse events
        - Treatment-related deaths

ICER's literature search and application of inclusion criteria resulted in the identification of six key studies, including one single-arm phase II trial of DARA and five phase III trials, i.e., one for each of the remaining five regimens of interest. Table 6 lists the six studies. Of note, there were no published studies of head-to head comparisons of the treatment regimens of interest.

[17] Midwest CEPAC. Treatment Options for Relapsed or Refractory Multiple Myeloma: Effectiveness, Value, and Value-Based Price Benchmarks Final Report and Meeting Summary. Boston, MA: ICER, June 2016.

Table 6. Key Trials of Multiple Myeloma Drug Regimens

Trial Name	Study Design	MM Drug Regimen	Comparator Regimen
ASPIRE	Open-label Phase III RCT	CFZ+LEN+DEX	LEN+DEX
SIRIUS	Open-label Phase II Single-arm Study	DARA	None
ELOQUENT-2	Open-label Phase III RCT	ELO+LEN+DEX	LEN+DEX
TOURMALINE-MM1	Double-blind Phase III RCT	IX+LEN+DEX	Placebo+LEN+DEX
PANORAMA-1	Double-blind Phase III RCT	PAN+BOR+DEX	Placebo+BOR+DEX
MM-003	Open-label Phase III RCT	POM+LoDEX	HiDEX

The comparative effectiveness review found the following regarding levels of evidence for each regimen:

Table 7. Level of Evidence for Multiple Myeloma Drug Regimens

Regimen	Evidence
CFZ+LEN+DEX	Moderate certainty of incremental or better net health benefit for both second- and third-line or subsequent therapy
ELO+LEN+DEX	
IX+LEN+DEX	
PAN+BOR+DEX	Insufficient evidence for second-line therapy; promising but inconclusive for third-line and subsequent therapy
POM+LoDEX	Insufficient evidence for second-line therapy; promising but inconclusive for third-line and subsequent therapy
DARA	Insufficient evidence to determine net health benefit for second- or third-line and subsequent therapy

For OS, final data were available from only two trials indicating a benefit associated with POM+LoDEX of 4.6 months of improved survival compared to hiDEX, and no statistically significant difference in OS for PAN+BOR+DEX compared to BOR+DEX. The review retrieved interim data showing an OS benefit with CFZ+LEN+DEX and ELO+LEN+DEX. For PFS, all regimens of interest evaluated with RCTs demonstrated statistically significant improvement in PFS compared to control treatment. The review also found similar rates of treatment-related deaths, discontinuation due to adverse events, and grade 3 or higher adverse events across all regimens except PAN+BOR+DEX.

Health-related quality of life data were available for four of the six regimens. CFZ+LEN+DEX: demonstrated greater improvements compared with LEN+DEX over 18 cycles of treatment. POM+LoDEX resulted in greater improvements in seven out of eight domains of health-related quality of life compared with high-dose DEX alone. There were no differences in health-related quality of life for ELO+LEN+DEX or IX+LEN+DEX compared with LEN+DEX.

ICER also conducted indirect comparisons of the regimens of interest on OS and PFS using techniques of Bayesian network meta-analysis. Findings from this additional analysis showed improvements in OS and PFS for each regimen versus its comparator, but no clinically meaningful differences between the regimens of interest were found. In its report, ICER indicated that there are limitations to its review, acknowledging that that the greatest source of uncertainty of the comparative net health benefit arises from the lack of truly comparative data across trials at the time of the assessment. ICER noted that many of the drugs were recently approved. In addition, the number of available studies and the absence of data for key subgroups precluded robust indirect comparisons of the regimens.

ICER worked with the University of Washington to conduct the cost-effectiveness analysis. The analysis found that all but one of the new regimens improved outcomes at substantial additional costs. Compared to LEN+DEX (baseline comparator), second- and third-line therapy with CFZ+LEN+DEX, ELO+LEN+DEX, and IX+LEN+DEX all resulted in incremental cost-effectiveness ratios higher than ICER's $100K-$150K/QALY threshold. According to ICER, "the cost per QALY that is generally accepted as 'reasonable' value in the U.S. is $50,000-$150,000 so CFZ+LEN+DEX, ELO+LEN+DEX, and IX+LEN+DEX at list prices would not represent good value in the long-term."

Table 8. Cost per QALY Gained for Second- and Third-line Treatments

Regimen	Cost per QALY as second-line treatment	Cost per QALY as third-line treatment
CFZ+LEN+DEX	$199,982	$238,560
ELO+LEN+DEX	$427,607	$481,244
IX+LEN+DEX	$433,794	$484,582

As a third-line therapy, PAN+BOR+DEX was estimated to provide more QALYs at a lower cost than LEN+DEX, with a cost per QALY of $10,230 (vs. BOR+DEX). However, ICER noted that the long-term cost-effectiveness remains uncertain because of concerns over the high rates of study discontinuation due to toxicity observed in the available literature.

POM+LoDEX and DARA were not included in the analysis because only single-arm data were available for DARA and therefore no incremental treatment effect vs. LEN+DEX could be estimated. Also, DARA and POM+LoDEX were studied in populations with more advanced disease (i.e., refractory to BOR and/or LEN), so the effects of these regimens were not considered comparable to those of the other regimens.

Based on the available evidence, members of the panel voted on the care value of regimens as second-line and third-line therapies. The majority of panel members determined each of the second-line therapies to be of intermediate value (Table 9). None voted high value for any of the regimens, but ICER reported that members who voted for intermediate value cited the significant clinical benefit of the regimens in spite of the finding that the cost-effectiveness of the regimens exceeded commonly cited thresholds.

Table 9. Care Value Vote for Second-line Treatments

Regimen	Low	Intermediate	High
CFZ+LEN+DEX	2 votes	9 votes	0 votes
ELO+LEN+DEX	4 votes	7 votes	0 votes
IX+LEN+DEX	4 votes	7 votes	0 votes

Table 10 presents the members' votes on the care value of third-line treatments. There were no high votes for the first three regimens. More panel members voted low value for these regimens as third-line treatments given that their incremental cost-effectiveness ratios were even higher for third-line treatment.

Table 10. Care Value Vote for Third-line Treatments

Regimen	Low	Intermediate	High
CFZ+LEN+DEX	2 votes	9 votes	0 votes
ELO+LEN+DEX	6 votes	5 votes	0 votes
IX+LEN+DEX	5 votes	6 votes	0 votes
PAN+BOR+DEX	4 votes	4 votes	3 votes

The next component of the framework was to determine potential short-term budget impact for these second-line and third-line therapies. None of the regimens surpassed the potential budget impact threshold of $904 million for a new drug. ICER estimated that approximately 33,900 patients in the U.S. would be eligible for second-line therapy. Assuming no coverage or reimbursement restrictions, the ICER analysis used an estimate of 75% of all eligible patients (25,455 patients) being prescribed CFZ+LEN+DEX, ELO+LEN+DEX, or IX+LEN+DEX over a five-year time horizon, with 25% of patients allotted to each of the three regimens. Based on these assumptions, ICER determined that the average potential budget impact per year would be approximately $226 million for CFZ+LEN+DEX, $395 million for ELO+LEN+DEX, and $330 million for IX+LEN+DEX over the five-year period.

Table 11. Potential Budget Impact (BI) for Second-line Regimens at 5 Years

Regimen	Number Treated*	Weighted BI per Patient	Average BI per year (millions)
CFZ+LEN+DEX	8,485	$133,907	$225.9
ELO+LEN+DEX	8,485	$232,848	$395.1
IX+LEN+DEX	8,485	$194,388	$329.9
Total	25,455	$186,777	$950.9

* At 75% uptake

ICER estimated that approximately 11,900 patients in the U.S. would be eligible for third-line therapy. Again, ICER applied assumptions of an unmanaged uptake rate of 75% (8,940 patients) for the third-line regimens over a period of five years, with 18.75% of patients allotted to each of the four third-line regimens. Based on these assumptions, the average potential budget impact per year was estimated to be approximately $59 million for CFZ+LEN+DEX, $99 million per year for ELO+LEN+DEX, $83 million for IX+LEN+DEX, and $11.8 million for PAN+BOR+DEX over the five-year period.

Table 12. Potential Budget Impact (BI) for Third-line Regimens at 5 Years

Regimen	Number Treated**	Weighted BI per Patient	Average BI per year (millions)
CFZ+LEN+DEX	2,235	$132,358	$59.2
ELO+LEN+DEX	2,235	$222,438	$99.4
IX+LEN+DEX	2,235	$185,379	$82.9
PAN+BOR+DEX*	2,235	$26,414	$11.8
Total	8,940	$141,648	$253.3

*Compared to BOR+DEX

** At 75% uptake

ICER also calculated value-based price benchmarks for each of the second-line and third-line regimens. ICER defines the benchmark price of a drug as the price range that would achieve cost-effectiveness ratios between $100K and $150K per QALY gained, subject to the assumed uptake rates, without exceeding the $904 budget impact threshold for new drugs. As noted above, none of the estimates for the potential budget impact of each of these regimens exceeded the budget impact threshold of $904 million when annualized over a five-year time horizon.

Estimated value-based price benchmarks for all of the second-line regimens, when compared to the lists prices, indicated that for each regimen, a discount from the WAC price would be required to achieve each incremental cost-effectiveness threshold.

Table 13. Value-Based Price Benchmarks for Second-line Regimens

Regimen	WAC Price per Vial/Capsule	Cost to Achieve $100K/QALY	Cost to Achieve $150K QALY	Value-Based Price Benchmark	Discount from List Price*
CFZ+LEN+DEX	$1,862	$673	$1,267	$673 to $1,267	32-64%
ELO+LEN+DEX	$2,358	$267	$588	$267 to $588	75-89%
IX+LEN+DEX	$2,890	$181	$587	$181 to $587	80-94%

*Discount from wholesale acquisition cost (WAC)

For third-line regimens, all estimated value-based price benchmarks, except for PAN+BOR+DEX, were lower than the WAC prices, again indicating that a discount from the WAC price would be required to achieve each incremental cost-effectiveness threshold. For third-line PAN, the value-based price benchmark was estimated to be $2,933 to $3,886 per capsule, which is substantially higher than the WAC price. These results reflect the results of the cost-effectiveness analysis, which yielded ratios greater than $150K/QALY for all of the third-line regimens except PAN+BOR-DEX, which the analysis showed to be cost-saving relative to BOR+DEX.

Table 14. Value-Based Price Benchmarks for Third-line Regimens

Regimen	WAC Price per Vial/Capsule	Cost to Achieve $100K/QALY	Cost to Achieve $150K/QALY	Value-Based Price Benchmark	Discount from List Price*
CFZ+LEN+DEX	$1,862	$432	$974	$432 to $974	48-77%
ELO+LEN+DEX	$2,368	$178	$466	$178 to $466	80-93%
IX+LEN+DEX	$2,890	$74	$440	$74 to $440	85-97%
PAN+BOR+DEX**	$1,222	$2,933	$3,886	$2,933 to $3,886	-

*Discount from wholesale acquisition cost (WAC)

**Compared to BOR+DEX

Since introducing its framework, ICER has incorporated value assessments it into eight of its posted completed reports. ICER is also one of the framework developers that actively seeks input and feedback from stakeholders through its processes for public comment on interventions being assessed and via its national call for suggestions for improving the framework. The national call that ICER issued in 2016 notes areas for potential revision, which appear to coincide with many of the aspects of the ICER framework for which various stakeholders have expressed concerns, including about the multiple myeloma report. These areas include: methods to integrate patient and clinician perspectives on the value of interventions, incremental cost-effectiveness ratios (appropriate thresholds, best practice in capturing health outcomes through the QALY or other

measures), methods to estimate market uptake and potential short-term budget impact of new interventions, and methods for setting thresholds for potential short-term budget impact to serve as useful alarm bells for policymakers to consider whether affordability may pose a substantial challenge.

Input and feedback from the national call will be used to inform a planned 2017 update (version 2.0) to ICER's framework. ICER also presented changes to the framework that it had already incorporated in June 2016, which ICER described as "version 1.5" of its framework.

B. Cross-Framework Comparisons of Multiple Myeloma Assessments

As is apparent from the descriptions above of the frameworks' assessments of treatments for multiple myeloma, there are certain similarities and marked differences among them. Among the differences is the selection of regimens. For the cross-framework comparisons, we identified regimens that were assessed by at least two of the value assessment frameworks/tools for the same diagnoses and compared their findings to the extent possible.

1. Active Multiple Myeloma Not Previously Treated

Two of the frameworks in this analysis examined treatments for active multiple myeloma that was not previously treated: ASCO and NCCN. Presented as an example of how it would apply its framework under development, ASCO's assessment focused on patients with advanced multiple myeloma who were not previously treated. The intervention of interest was the drug regimen of bortezomib, melphalan, and prednisone (BOR+MEL+PRED) relative to comparator (MEL+PRED). This can be compared to NCCN's guidelines with Evidence Blocks for multiple myeloma, which lists the combination of BOR+MEL+PRED as a preferred Category 1 regimen for the primary treatment of multiple myeloma, specifically for non-transplant candidates. Note that the NCCN Evidence Blocks are designed to present an absolute assessment, rather than a relative assessment. ASCO and NCCN based their assessments on the same landmark RCT, the VISTA trial, which compared BOR+MEL+PRED to MEL+PRED (control) in patients who were not candidates for autologous stem cell transplantation.

Table 15. Comparison of ASCO and NCCN Findings for BOR+MEL+PRED*

	ASCO Assessment Findings*	NCCN Evidence Blocks Findings
Intended Audience(s)	• Clinicians, patients	• Clinicians, patients

Primary Output(s)	• NHB score of 47 out of possible 130 points
	  • Clinical benefit score: 32
	  • Toxicity score: 0
	  • Bonus points: 15 (improved treatment-free survival)
	• Average monthly cost: $7,042
		• Efficacy of regimen: 4 (very effective)
		• Safety of regimen: 3 (mildly toxic)
		• Quality of evidence: 4 (good quality)
		• Consistency of evidence: 4 (mainly consistent)
		• Affordability of regimen: 3 (modestly expensive)

Evidence/ Data Sources	• VISTA Trial	• VISTA Trial
Evidence Synthesis/ Rating	• N/A	• Quality of evidence: 4 (good quality) • Consistency of evidence: 4 (mainly consistent) • Category I

*ASCO NHB is relative to comparator: MEL+PRED. NCCN scores are not relative to comparator.

Although each tool reports different types of findings, because both tools are based on the same evidence in the instance of this comparison, their findings appear to be aligned in certain ways. At a more general level, both ASCO and NCCN focused the development of their tools on helping patients and providers with the decision-making process and cover similar domains that contribute to the results of each tool.

Compared to the control regimen of MEL+PRED, ASCO assigned the BOR+MEL+PRED regimen a clinical benefit score of 32 points plus 15 bonus points, based on OS and treatment-free survival, demonstrating that the regimen is more effective than the control. Similarly, in terms of efficacy, NCCN Evidence Blocks scored the regimen a 4 for very effective, which NCCN defines as a regimen that often provides long-term survival advantage or has curative potential.

Although the regimen was associated with 42 grade 3-5 toxicities (versus 35 grade 3-5 toxicities for the control), ASCO neither awarded nor deducted points from the regimen for toxicity, indicating little net difference in the high-severity toxicities associated with the regimen relative to the control. NCCN rated the safety of the regimen a 3 for mildly toxic, which may reflect the extent of toxicities associated with the regimen.

While NCCN also rated the quality and consistency of the evidence as 4, i.e., for good quality and mainly consistent evidence, ASCO does not provide a comparable rating of the quality or consistency of the evidence.

ASCO reported an average monthly cost for the regimen of $7,042 (versus $279 per month for the control regimen). However, the cost used by ASCO is the drug acquisition cost only and does not report the patient copayment or account for non-drug costs. NCCN states that its affordability score reflects the overall total cost of a therapy, including but not limited to acquisition, administration, in-patient vs. out-patient care, supportive care, infusions, toxicity monitoring, antiemetics, growth factors, and hospitalization. As NCCN uses average panel scores that align with undefined categories for affordability (e.g., very inexpensive), its affordability score and ASCO's cost data cannot readily be compared.

2. Refractory or Relapsed Multiple Myeloma

Three of the value assessment frameworks examined regimens for refractory or relapsed multiple myeloma, including DrugAbacus, NCCN, and ICER. Table 16 lists seven regimens for refractory or relapsed multiple myeloma that were common across at least two of those three frameworks. To illustrate the different inputs and outputs of the frameworks, assessments for regimen 1 in Table 16, which is addressed by DrugAbacus and NCCN, and regimen 6 in Table 16, which is addressed by all three of the frameworks, are described below.

Table 16. Regimens Assessed by At Least Two Frameworks

	Regimen	DrugAbacus	NCCN	ICER
1	Bortezomib (BOR)	X	X	
2	Carfilzomib + lenalidomide + dexamethasone (CFZ+LEN+DEX)		X	X
3	Elotuzumab + lenalidomide + dexamethasone (ELO+LEN+DEX)		X	X
4	Ixazomib + lenalidomide + dexamethasone (IX+LEN+DEX)		X	X
5	Daratumumab monotherapy (DARA)		X	X

	Regimen	DrugAbacus	NCCN	ICER
6	Panobinostat + bortezomib + dexamethasone (PAN+BOR+DEX)	X	X	X
7	Pomalidomide + low-dose dexamethasone (POM+LoDEX)	X	X	X

a. BOR

DrugAbacus and NCCN assessed bortezomib (BOR) as a treatment for relapsed/refractory multiple myeloma. As described above, DrugAbacus uses clinical trial data from a drug's FDA approval file for the drug's first indication. However, the tool does not specify the trial data that were used. The frequently asked questions page of the DrugAbacus website mentions a paper by "Howard et al." as a key data source. Further inspection led to the retrieval of an article with Howard as first author that includes the relevant clinical trial information for BOR; we assumed that this was used to populate DrugAbacus.[18] The appendix referenced the Assessment of Proteasome Inhibition for Extending Remissions (APEX) trial, an RCT that compared BOR to high-dose DEX in patients with relapsed multiple myeloma.[19] This trial is the same trial that NCCN examined for its assessment of BOR.

The DrugAbacus output is an Abacus-estimated monthly price for a drug. The price generated by DrugAbacus depends on the user's preferences for each of the eight domains that comprise the tool. The tool also shows the actual monthly price of a drug (as paid by Medicare). Varying the values of each of the domains to their extremes yielded an estimated monthly DrugAbacus price ranging from $841 to $728,361. However, a more useful estimate may be the tool's estimate of $9,442, which derives from setting an extra year of life at $132,000, incorporating a toxicity discount of 15%, and assigning no weight to the additional parameters. DrugAbacus reports that the actual monthly cost for BOR is $4,474. As noted above, the efficacy and safety values for drugs included in the tool, including BOR, which are used to help calculate the Abacus price, exist within the DrugAbacus tool but are not readily accessible or visible to the user; neither is how these data were estimated.

Table 17. Comparison of DrugAbacus and NCCN Assessment for BOR

	DrugAbacus Findings	NCCN Evidence Blocks Findings
Primary Intended Audience(s)	• Policymakers, payers, industry	• Clinicians, patients
Primary Output(s)	• Minimum estimated monthly Abacus price: $841 • Maximum estimated monthly Abacus price: $728,361 • Estimated monthly Abacus price at $132,000 per life year and 15% toxicity discount: $9,442 • Actual monthly cost: $4,474	• Efficacy of regimen: 3 (moderately effective) • Safety of regimen: 4 (occasionally toxic) • Quality of evidence: 4 (good quality) • Consistency of evidence: 4 (mainly consistent) • Affordability of regimen: 2 (expensive)
Evidence/Data Sources	• APEX Trial	• APEX Trial

[18] Howard DH et al. 2015.

[19] Richardson PG, Sonneveld P, Schuster MW, Irwin D, et al. Bortezomib or high-dose dexamethasone for relapsed multiple myeloma. N Engl J Med. 2005 Jun 16;352(24):2487-98.

	DrugAbacus Findings	NCCN Evidence Blocks Findings
Evidence Synthesis/Rating	• N/A	• Quality of evidence: 4 (good quality) • Consistency of evidence: 4 (mainly consistent)

The NCCN Evidence Blocks focus on the efficacy and safety of BOR and the quality and consistency of the available evidence. NCCN scored the efficacy of BOR a 3 for moderately effective. NCCN panel members rated the safety of BOR as occasionally toxic, which means BOR was associated with rare significant toxicities or low-grade toxicities only.

NCCN panel members also rated BOR's affordability as expensive. However, this rating cannot be readily compared to DrugAbacus' results, for reasons similar to those described above in the comparison of ASCO and NCCN findings for BOR+MEL+PRED.

b. PAN+BOR+DEX

DrugAbacus, NCCN, and ICER all included the regimen PAN+BOR+DEX. Panobinostat is an oral agent that was FDA-approved in 2015 as the first histone deacetylase (HDAC) inhibitor for the treatment of multiple myeloma. It is indicated for use in combination with bortezomib and dexamethasone in patients who have received at least two prior lines of treatment, including bortezomib and an immunomodulatory drug (IMiD).

Of these three assessment frameworks that included PAN+BOR+DEX, only NCCN's Evidence Blocks is intended for use by providers and patients; DrugAbacus focuses on policymakers, and ICER focuses on policymakers and payers. Note that ICER's assessment is relative to a comparator, while NCCN and DrugAbacus' assessments are absolute. Table 18 compares the three frameworks' assessments of PAN+BOR-DEX, and presents the very different outputs from each.

Table 18. Comparison of DrugAbacus, NCCN, and ICER for the Assessment for PAN+BOR+DEX

	DrugAbacus Findings	NCCN Evidence Blocks Findings	ICER Findings
Primary Intended Audience(s)	• Policymakers, payers, industry	• Clinicians, patients	• Payers, policymakers, industry
Primary Output(s)	• Minimum estimated monthly Abacus price: $661 • Maximum estimated monthly Abacus price: $169,006 • Estimated monthly Abacus price at $132,000 per life year and 15% toxicity discount: $7,817 • Actual monthly cost: $10,625	• Efficacy of regimen: 3 (moderately effective) • Safety of regimen: 2 (moderately toxic) • Quality of Evidence: 4 (good quality) • Consistency of evidence: 4 (mainly consistent) • Affordability of regimen: 2 (expensive)	• Median PFS: 12.5 mo vs 4.7 mo; statistically significant difference* • Relative to other regimens, presented a more severe toxicity profile • CE: $10,230 per QALY** • Care Value votes: Low (4 votes), Intermediate (4 votes), High (3 votes) • Weighted BI per patient:*** $26,414 • Avg. BI/yr (millions): $11.8 • WAC price per vial/capsule: $1,222 • Value-based price benchmark: $2,933 to $3,886; no discount from WAC needed to achieve CE
Evidence/ Data Sources	• PANORAMA-1 Trial (Phase III)	• PANORAMA-1 Trial (Phase III) • PANORAMA-2 Trial (Phase II trial)	• PANORAMA-1 Trial (Phase III)
Evidence Synthesis/ Rating	• N/A	• Quality of Evidence: 4 (Good quality) • Consistency of Evidence: 4 (Mainly consistent) • Category 1	• Insufficient evidence for second-line therapy; promising but inconclusive for third-line and subsequent therapy

* Based on subgroup analysis of 147 patients with multiple myeloma who had received at least 2 prior treatments, including BOR and an IMiD in the phase III PANORAMA-1 trial.
** Given lingering concerns over high rates of study discontinuation due to toxicity, the long-term cost-effectiveness remains uncertain.
*** For 5-yr horizon, drug costs and cost offsets apportioned assuming 20% of patients in uptake target initiate therapy each yr.

Regarding the evidence used, it appears that all three frameworks relied on the PANORAMA-1 Trial, which was used to obtain FDA approval for the drug. The NCCN guidelines described both the PANORAMA-2 (phase II) and PANORAMA-1 Trial (phase III) in the discussion section, assigning the regimen a Category 1 for evidence and consensus, which, again, indicates it is based on high-level evidence and uniform NCCN consensus that the intervention is appropriate. NCCN assigned a score of 4 to each to the Evidence Blocks for quality of evidence and consistency of evidence. ICER identified the PANORAMA-2 Trial (phase II) and PANORAMA-1 Trial (phase III) as the sources of data for PAN in its report but rated the available evidence as insufficient for second-line therapy and promising, but inconclusive, for

third-line subsequent therapy. As previously described, there are no references for the data sources used to populate the DrugAbacus tool. The paper by Howard et al. did not present the relevant information. This analysis assumes DrugAbacus also used PANORAMA-1 Trial.

The three frameworks present different value outputs. As noted above, assuming a dollar value of $132,000 per life year and toxicity discount of 15% in the tool yields a DrugAbacus price of $7,817 per month, which is less than the actual monthly cost of $10,625. ICER estimated a value-based price benchmark of $2,933 to $3,886 per vial/capsule. Compared to the WAC price per vial/capsule of $1,222, ICER determined that there would be no discount needed from the WAC price to achieve cost-effectiveness. For PAN, NCCN's panel scored the affordability Evidence Block a 2, qualitatively categorizing it as an "expensive" regimen. None of these various framework cost measures can readily be compared.

V. Discussion

A cursory examination of the four frameworks' assessments reveals four different-looking sets of findings of varying relevance to certain users and other stakeholders. As noted above, this variation raises ambiguity and poses confusion for users. Apart from the specific aspects of treatments for multiple myeloma, the frameworks have fundamental differences in their purposes, target audiences, methodologies, provisions for stakeholder input, and user interaction.

These frameworks yielded a range of findings reflecting differences in patient populations and drug regimens assessed for multiple myeloma. NCCN assessed treatments for the widest range of types of the disease, including active multiple myeloma that has not been previously treated and relapsed and refractory multiple myeloma. ASCO focused on one treatment for advanced multiple myeloma. DrugAbacus included three drugs for relapsed or refractory multiple myeloma. ICER selected regimens based on FDA-labeled indications for treatment of relapsed or refractory multiple myeloma, as well as treatment approaches that are currently of greatest clinical interest, based on stakeholder input.

The four frameworks draw on expertise in different ways, ranging from a small set of in-house researchers for DrugAbacus to national expert panels comprised largely of oncologists and oncology researchers (and some patient representation) for each type of cancer for NCCN. ICER uses combinations of staff and external multistakeholder panels of ICER, although these groups generally do not include topic-specific clinical experts. However, ICER does engage topic-specific clinical experts in developing the draft scope of its reports as expert report consultants and at policy roundtables. ICER's internal multidisciplinary team is complemented by academic-based teams that conduct much of ICER's economic modeling.

While certain components of value are shared across most or all of these frameworks, including consideration of efficacy and toxicity and some form of costs/affordability, other value components are different, such as several of the DrugAbacus domains (e.g., novelty, R&D costs, rarity, and unmet need); ICER's qualitative assessment of other benefits and harms and contextual considerations; and ASCO's consideration of quality of life, tail of the curve survival, treatment-free interval, and symptom palliation. Also, the frameworks differ in the extent to which they integrate their respective value components or domains.

Using the original ASCO framework, ASCO presented information about a single drug regimen versus a comparator for multiple myeloma, based largely on a single RCT, in the form of NHB (net health benefit) score, which is a function of clinical benefits and toxicity as well as symptom palliation and treatment-free survival. This is accompanied by drug acquisition cost, which may have limited relevance to patients' actual out-of-pocket costs. The framework can serve as a tool or basis for further discussion by clinicians and patients.

The NCCN framework, in the form of its Evidence Blocks, is also designed to support clinician and patient decision-making. The Evidence Blocks—one for each drug regimen—accompany the relevant NCCN practice guideline for different forms of multiple myeloma, based on literature and expert consensus. This combination of information may be more intuitive to patients, given that each Evidence Block comprises five domains, each of which is kept discrete and rated on a 1-5 visual scale. However, the affordability domain is a broad cost construct that may have only limited relevance to patients. As such, accompanying the NCCN multiple myeloma guideline, the Evidence Blocks provide a decision support tool for clinicians and patients that can complement the guidelines.

While DrugAbacus includes drug-specific estimates for efficacy, tolerability, novelty, research and development costs, rarity, population burden, unmet need, and prognosis, these estimates are not visible to the user. These estimates populate an equation and interact with user-selected weights for the eight domains to generate a value-based monthly price for each of three multiple myeloma drugs. This value-based monthly price is compared to the actual monthly price to Medicare for each drug. The eight parameters are among those considered by policy-makers, industry, and others, and others as contributing, or being otherwise relevant, to market prices. Several of these are not relevant to clinical decision-making for individual patients. Rather than being a tool for clinicians and patients, DrugAbacus is directed primarily to policymakers, payers, and industry, enabling them to test how varying the relative weights of the eight parameters would affect a value-based drug price, and how such a price compares to an actual market price.

In addition to results pertaining to comparative clinical effectiveness (efficacy and safety), the ICER framework, directed primarily to payers, policymakers, and industry, generates results that include economic information that is not produced by the other frameworks for the treatment regimens. This includes long-term cost-effectiveness ratios (in the form of cost per QALY gained) and short-term potential budget impact information. Further, ICER generates value-based price benchmarks that would achieve particular levels of long-term cost-effectiveness and remain below potential levels of short-term budget impact.

A. Opportunities for Improving Quality and Utility of Frameworks

Although differences across the frameworks are to be expected given their fundamental differences in purpose, target audiences, and methodologies, their application to multiple myeloma points to opportunities for developers and stakeholders to work towards some common elements or guiding practices to improve their quality and utility to their intended users. Among these opportunities are the following.

1. Evidence Sources and Quality

Framework developers should carefully consider, and be fully transparent regarding, their methods, sources, and criteria for selection of evidence. Further, the scope and limitations of the evidence sources, e.g., inclusion and exclusion criteria, gaps in the evidence, and how those limitations affect assessment findings, should be addressed. The frameworks should rate the quality of the evidence used in their assessments in a transparent manner using standard, accepted methods.

Across these frameworks, the primary evidence sources for an assessment range from reliance on single RCTs to bodies of evidence comprising RCTs, other clinical trials, observational studies, conference abstracts, regulatory review content, and network meta-analyses. Methods for evaluating the quality of the evidence vary; in some cases, there is limited or no evaluation of the quality of the evidence.

Table 19. Level of Evidence Assessed by Frameworks

Framework	Types of Evidence Used	Evidence for Multiple Myeloma Regimens in this Analysis	Evaluated for Quality
ASCO	Single RCT (usually landmark/pivotal trial)	VISTA Trial for BOR+MEL+PRED	No
Drug Abacus	Single RCT (used to gain FDA approval for first indication)	• APEX Trial for BOR • MM-003 Trial for POM+LoDEX • PANORAMA Trial for PAN+BOR+DEX	Partial*
NCCN	Broad evidence base (e.g., RCTs, non-RCTs, meta-analysis or systematic reviews, clinical case reports, case series), clinical experience	• Various studies including: VISTA for BOR+MEL+PRED APEX for BOR ASPIRE for CFZ+LEN+DEX SIRIUS for DARA ELOQUENT-2 for ELO+LEN+DEX TOURMALINE-MM1 for IX+LEN+DEX PANORAMA-1 for PAN+BOR+DEX MM-003 for POM+LoDEX • In some cases, the only available published evidence was based on a trial conducted for FDA approval/landmark trial (e.g., VISTA, APEX, PANORAMA)	Yes
ICER	Publicly available, peer-reviewed literature on clinical and cost-effectiveness, grey literature	Six key studies: ASPIRE for CFZ+LEN+DEX SIRIUS for DARA ELOQUENT-2 for ELO+LEN+DEX TOURMALINE-MM1 for IX+LEN+DEX PANORAMA-1 for PAN+BOR+DEX MM-003 for POM+LoDEX	Yes

* Regarding quality of evidence, the DrugAbacus website states only that a "level-of-evidence grade is applied to the measure of overall survival benefit, such that two drugs with equivalent trial results will receive different efficacy grades if the trial for one drug was of higher quality."

DrugAbacus considers single clinical trials, as does the current ASCO methodology. DrugAbacus uses the clinical trials that were conducted to gain FDA approval for the drugs. This

was the case for each of the three drugs that appear in the tool when the user selects "myeloma" as the condition of interest. DrugAbacus applies a level-of-evidence grade to the measure of OS benefit in each trial, such that two drugs that have equivalent trial results will receive different efficacy grades if the trial for one drug was of higher quality. ASCO examined only one regimen for the primary treatment of advanced multiple myeloma in patients who were not previously treated and who were not transplant candidates. ASCO's developers have indicated that its revised framework will consider the broader evidence base for treatments, not just single clinical trials. ASCO should also consider incorporating an assessment of the quality of the evidence using validated methods.

While NCCN and ICER examine a far broader evidence base than ASCO and DrugAbacus, they do so differently. ICER conducts a systematic literature review using standard methods. NCCN conducts an extensive literature search and considers a wide range of available evidence, but also relies on its expert panel members' knowledge of the evidence and their clinical experience. The available evidence for certain newer regimens for multiple myeloma was very limited. In some instances, there was only one clinical trial available at the time NCCN assessed this topic. This was the case for the two examples described above, i.e., BOR+MEL+PRED, which was common to the ASCO and NCCN assessments for primary treatment of multiple myeloma (VISTA trial), and PAN+BOR+DEX, which was one of the regimens common to DrugAbacus, NCCN, and ICER for treatment of refractory or relapsed multiple myeloma (PANORAMA trial).

Where head-to-head trials are few or unavailable, ICER may use indirect treatment comparisons. However, differences in trial design, such as patient inclusion and exclusion criteria, how outcomes of interest are defined, and other differences among trials can introduce bias into indirect comparisons. For refractory or relapsed multiple myeloma, ICER conducted indirect comparisons of the regimens of interest on OS and PFS using techniques of Bayesian network meta-analysis. ICER acknowledged that, given the lack of truly comparative data across trials, the certainty of the comparative net health benefits of the regimens examined is very limited. The limited number of available studies as well as the absence of data for certain key subgroups precluded ICER from conducting robust indirect comparisons of the regimens in its review. This highlights the necessity of transparency regarding the limitations of available evidence and methods for drawing comparisons, as well as the reliability of the resulting findings. Further, this underlines the need for explicit provisions for updating assessments if and when better evidence emerges, including from relevant direct comparative trials.

2. Clinical and Economic Outcomes

Attempting to compare or align the various clinical and cost outputs from the frameworks is difficult and may not be useful. As shown in some of the examples above, even where the frameworks focused on the same patient population and regimen(s) and used the same evidence, clinical and cost outputs varied across frameworks.

Table 20 shows overlap in the types of clinical inputs considered by each of the guidelines. ASCO, DrugAbacus, and NCCN value and/or give greater weight to OS compared to PFS. However, in the case of multiple myeloma, some stakeholders contend that OS is rarely the basis for drug approval and may no longer be a reliable endpoint in myeloma given the availability of new agents for relapse, and therefore that PFS is more useful.

Table 20. Clinical Inputs Used by Frameworks

Clinical Evidence	ASCO	DrugAbacus	NCCN	ICER
Overall survival*	X	X	X	X
Progression-free survival	X	X	X	X
Overall response rate	X	X	X	X
Treatment-free interval	X			
Toxicities/adverse events	X	X	X	X
QoL/palliation	X			X
Disease burden		X		X
Unmet need		X		

*ASCO, DrugAbacus, and NCCN value/give greater weight to OS than to PFS.

Table 21 presents the various types of cost-related components of each of the frameworks. There is minimal overlap of these across the frameworks.

Table 21. Cost/Affordability Components of Frameworks

	Cost / Affordability	Source(s)
ASCO	Drug acquisition cost per month	Average sales price as of October 2014 for intravenous therapies and information from UHC for oral drugs
Drug Abacus	• DrugAbacus price • Average monthly cost	• Calculated using DrugAbacus equation • Based on Medicare payment
NCCN	Affordability Evidence Block score	Expert panel members' knowledge of total costs related to use of drug
ICER	• Wholesale Acquisition Cost (WAC) price	• Manufacturer's published catalog or list price for a drug product to wholesalers as reported to First Databank by the manufacturer. WAC does not represent actual transaction prices and does not include prompt pay or other discounts, rebates, or reductions in price.
	• Cost/QALY gained	• Calculated using CE model
	• Potential budget impact	• Assumed uptake rates times WAC
	• Value-based price benchmarks	• Price benchmarks to show price at which (1) drug reaches $100K/$150K per QALY gained and (2) potential budget impact reaches $904M

Stakeholders have also called for inclusion of more patient-centered outcomes, indicating that assessments of multiple myeloma treatments to date lack consideration of patient preferences and experiences. Concerns have focused on the lack of factors important to patients, including health-related QoL, ease of use, management of toxicities and side effects (including low-grade, chronic side effects), and financial toxicity (i.e., patient cost burden and its implications).

In response to these concerns, framework developers increasingly recognize the importance of these outcomes and have expressed their interest in considering them. ASCO has indicated that,

once there are credible tools for patient-reported outcomes and more trials include them, ASCO will incorporate them into its value assessment framework. ICER indicated that it contacted patient groups and advocacy organizations to gain input on the most important outcomes, other benefits and disadvantages of new treatment options, and current context considerations for management of multiple myeloma. ICER stated that it updated its final report to clarify these considerations where relevant. ICER also reported adding further context to its report for other dimensions of care that are important to patients, including valuation of quality of life over its span and consideration of low-grade, chronic side effects.

3. Timing of Assessments and Need for Updates

For conditions such as multiple myeloma, where new treatments are emerging, framework developers need to consider the tradeoffs of user demand for timely findings and sufficiency of evidence for credible findings. In the case of multiple myeloma, stakeholders have pointed to the limited evidence base for these treatments at the time the assessments were conducted and called for assessments to be conducted later when more evidence would have been available. Although NCCN and ICER looked more broadly for available evidence than other frameworks, only one clinical trial was available for some treatments.

Assessments conducted at or near the time of market approval would not include subsequent evidence that might alter the understood value of treatment regimens. Still, for NCCN and ICER, conducting assessments at or near the time of regulatory approval of a new cancer drug responds to the decision-making demands of their main users, who are largely physicians and patients for NCCN and payers for ICER. While more evidence is likely to emerge over time, many users still must make practical decisions about newly available cancer drugs.

Additional useful evidence might accrue from post-marketing trials and various sources of real-world evidence, including on patient-centered outcomes and low-grade toxicities that are not routinely available in clinical trials conducted for regulatory approval. Frameworks should have provisions for prompting assessment updates, whether periodically or upon availability of relevant new evidence. Of the four frameworks that assessed treatments for multiple myeloma, only NCCN has a process for updating its guidelines with Evidence Blocks with any new evidence on an annual basis. NCCN will also look at new evidence between these annual reviews if it receives a request to do so and will make revisions as needed. Although ICER recognizes the need for more evidence on multiple myeloma treatments and that further evidence is likely to emerge, ICER has noted that its assessments are largely one-time efforts, and that it does not plan to review or update assessments at regular intervals. ICER would benefit its users by conducting updates of topics for which treatment protocols are evolving. As ASCO further develops its methodology, it should consider provisions for updating its value assessments upon appearance of relevant new evidence. DrugAbacus does not update the data in its tool. Recently, it added indication-based pricing for four drugs. It may want to consider expanding that feature to more drugs and updating the data on a regular basis.

4. Need for Greater Transparency

The wide variation in the outputs of these frameworks' assessments of treatments for multiple myeloma raises ambiguity and poses confusion for users. Users that are unable to, or uninterested in, discerning the intended uses and underlying assumptions of these frameworks

may misinterpret or misapply their results. Limitations in transparency of evidence sources and methods weaken the credibility, reliability, and/or utility of certain aspects of frameworks and their findings/results.

Certainly, transparency affects the reproducibility of the results of the assessments of these frameworks. For example, the ASCO framework is largely transparent. It showed how data from the designated clinical trial was used to determine the NHB for BOR+MEL+PRED. In contrast, DrugAbacus has limited transparency, as noted above. Although it provides the general equation it uses to determine the DrugAbacus price, it is difficult to reproduce the Abacus price for each drug. There were no references to the specific trials it used, nor does the tool readily make accessible the data from these trials that were entered into the equation for each multiple myeloma drug included in the tool. The tool does not specify what type of multiple myeloma these drugs are used to treat. With a separate, external search for the first FDA-approved indications, one can identify the type of multiple myeloma these drugs were approved to treat.

Although ICER's overall transparency is generally favorable, there are opportunities for improved transparency. For example, ICER has extensive and clear provisions for soliciting comments, but there is less clarity regarding how and to what extent it addresses these comments in its assessments. Some stakeholders have also expressed concern that ICER's economic modeling is insufficiently transparent and have been unable to replicate the model results. ICER indicates that its modeling analysis plans are intended to provide sufficient information for experienced researchers to be able to replicate the economic model and analyses. However, in order to protect the intellectual property rights of ICER's external collaborators, the actual executable models and associated computer code are not provided as part of the deliverable to ICER. To help ensure the credibility of these models, ICER and its collaborators could consider arranging for outside experts to examine or test these models in a way that would not compromise their intellectual property.

Notwithstanding their respective current levels of transparency, more would enable a broader group of users and other stakeholders to perceive the relative merits and utility of these frameworks' assessments of this clinical topic as well as others. In particular, frameworks should be more explicit and otherwise transparent about:

- Intended purpose(s)
- Intended users/audience
- Respective roles and responsibilities of framework staff, advisory panels, consultants, other contractors, and sponsors
- Selection of interventions and patient populations
- Sources of evidence
- Inclusion and exclusion criteria for evidence
- Solicitation and incorporation of problem formulation, data and evidence, and perspectives of patients and expert stakeholders
- Data entry and use in scoring, equations, algorithms, and models

- Limitations of frameworks and their outputs
- Guidance on appropriate use/applications of frameworks and their findings

B. Value Assessment Frameworks are Here to Stay

Value assessment frameworks are still works in progress. Even so, the four addressed in this paper are underway in the policy arena and, in different ways, already influencing decision-making. The factors that have fueled the demand for these frameworks are not subsiding.

Examining how these four value frameworks assess the same condition, multiple myeloma, helps to illuminate their respective methodologies as well as their strengths and areas for improvement. Delving into these differences helps to explain how it is that these frameworks, all of which have clinical and economic components and draw from overlapping evidence sources, could yield such different findings about the value of interventions for multiple myeloma. These different findings are intended to address the information needs of different sets of target users. As such, there is no single right set of answers or findings across these frameworks.

As this analysis also reveals, there are clear opportunities for each framework to deliver findings that are more transparent, rigorous, current, and practically relevant for their respective target users. These opportunities are readily available to the developers of these frameworks, all of whom have indicated an interest in, and to varying extents have followed through on, pursuing such improvements.