ADaM Design and Programming Standardization for Oncology Efficacy Endpoints Based on RECIST 1.1

Ellen Lin, Matt Ness, and Yinghui Wang, Seagen Inc.

ABSTRACT

RECIST 1.1 tumor response criteria are commonly used in clinical trials to evaluate treatment response of solid tumors, such as breast and colorectal cancer. Common oncology efficacy endpoints, such as the best overall response (BOR), duration of response (DOR), and progression-free survival (PFS), are often based on RECIST 1.1 and usually involve complex data collection and derivations. While sponsors strive to make such collection and derivation consistent across study protocols, there are often study-specific considerations that further challenge the programming of these efficacy endpoints.

In this paper, we will share an innovative, intuitive, highly structured ADaM design Seagen used successfully in several recent filings to support the analyses of such common oncology efficacy endpoints based on RECIST 1.1. We will explain how we standardize these endpoint derivations into parameterized sections and data sets that build upon one another in an easily understood progressive data flow to significantly improve programming efficiency and quality. With the resulting analysis-ready ADaM structures and step-by-step data flows, we have succeeded in separating the study-specific from the common derivations, simplified the programming of these complex endpoints, and created opportunities to develop standards-driven departmental macros to further elevate speed and quality for several components embedded within this design.

1. INTRODUCTION

Assessment of the change in tumor burden is an important feature of the clinical evaluation of cancer therapeutics; both tumor shrinkage and disease progression are useful endpoints in oncology clinical trials. The Response Evaluation Criteria in Solid Tumors (RECIST) is a standard way to measure how well a cancer patient responds to treatment. RECIST-based statistical analysis endpoints involve complex data collection and derivations that pose challenges for statistical programming.

Section 2 and 3 give an overview of the commonly used RECIST version 1.1 (RECIST 1.1) and related endpoints; Section 4 details pertinent ADaM designs that Seagen implements and provides some hypothetical data examples that illustrate data structures and key derivations; and Section 5 describes examples where programming of complex algorithms can be standardized as departmental macros following the standard ADaM models.

2. RECIST 1.1 AT A GLANCE

RECIST was originally published in 2000 and revised (i.e., RECIST 1.1) in 2009. RECIST allows for uniform reporting of imaging and reproducible analysis. It is the most widely used set of criteria for assessing a patient's response to treatment in solid tumors. Investigators, cooperative groups, industry, and government authorities have adopted these criteria in the assessment of treatment outcomes.

To use RECIST 1.1, a clinical study collects the following tumor assessments typically done by X-rays, CT scans, or MRI scans:

- At study baseline or any re-baseline, target lesions and non-target lesions are identified and target lesions are measured. The measurements include the longest diameter of non-nodal target lesions or the short axis of nodal target lesions, and the sum of diameters (SoD) of the target lesions.

- At each planned or unscheduled post-baseline time point the measurement of each target lesion is repeated, the status of each non-target lesion is determined, and an examination for any equivocal or unequivocal new lesions is performed.

- At each planned or unscheduled post-baseline time point the target and non-target lesion responses are evaluated, any new lesion progression is assessed, and the overall response is determined.

Lesion evaluation methodology and the overall response algorithm is specified in the RECIST 1.1 guideline.

Overall response is classified into complete response (CR), partial response (PR), stable disease (SD) for patients with target lesions at baseline, non-complete response/non-progressive disease (non-CR/non-PD) for patients with only non-target lesions at baseline, and progressive disease (PD). To simplify the concepts, SD and non-CR/non-PD will both be referred as SD in the rest of this paper.

The timepoint (i.e., by-visit) data including the derived parameters of SoD and overall response are usually reported by investigators (INV). For clinical studies where objective response (i.e., CR or PR) is the primary endpoint, and in particular where key drug development decisions are based on the observation of a minimum number of responders, it is recommended that all claimed responses be reviewed by independent experts we'll call Blinded Independent Central Review (BICR). If the study is a randomized trial, BICR reviewers should be blinded to treatment assignment. Data adjudicated and provided by BICR would be processed and analyzed separately from INV data.

Analysis endpoints categorized by the RECIST criteria often include best overall response (BOR), objective response rate (ORR), clinical benefit rate (CBR), duration of response (DOR), and progression-free survival (PFS); these have been used as the primary or supportive data for regulatory approvals by the FDA and EMA.

3. EFFICACY ENDPOINTS BASED ON RECIST 1.1

There are two types of endpoints, categorical (including binary) endpoints for evaluating tumor shrinkage and response rate, and time-to-event (TTE) endpoints for the calculation of response duration and progression-free survival, that have been commonly adopted in oncology trials.

In a simple study design with one treatment period, these endpoints summarize treatment effect across all protocol-specified time points. It is often also required to only include time points prior to the start of any subsequent cancer-targeting therapy when deriving BOR based categorical endpoints. This requirement is incorporated in the definitions and derivations discussed throughout the rest of this paper.

3.1 CATEGORICAL ENDPOINTS

BOR is the best response recorded across all post-baseline time points prior to the start of any subsequent cancer-related therapy (i.e., on or before the new therapy start date). Depending on the nature of the study and the protocol requirements, confirmation of CR or PR (e.g., a CR or PR confirmed at a subsequent assessment at least 4 weeks later) may or may not be required. Also, when SD is believed to be the best response, it must meet the protocol-specified minimum time from baseline (e.g., 5 weeks from the randomization date). BOR is a categorical endpoint defined by RECIST 1.1.

Related to BOR, two types of binary endpoints will also be referenced in this paper:

- **ORR** (objective response rate) is defined as the proportion of subjects whose BOR is CR or PR.

- **CBR** (clinical benefit rate) is defined as the proportion of subjects whose BOR is an objective response (CR or PR) or who achieve a sustained SD (i.e., an SD that lasts for at least a pre-specified duration of time before the occurrence of PD or death). The calculation of SD duration involves similar censoring rules as those for PFS that are described in Table 1 of Section 3.2.

In the rest of this paper, we will borrow the abbreviation BOR for the parameter code of best response, ORR for the parameter code of objective response, and CBR for the parameter code of clinical benefit.

3.2 TIME-TO-EVENT ENDPOINTS

Time-to-event endpoints typically include DOR, PFS, and any sensitivity analysis versions of DOR or PFS. A PFS sensitivity analysis example will be given at the end of this section. Overall survival (OS) will not be discussed in this paper as it is not a RECIST 1.1-based endpoint.

DOR is measured from the first documented objective response (CR or PR) to the first documented PD or death due to any cause, whichever comes first. DOR is only calculated for the subgroup of subjects

achieving a CR or PR. Censoring rules for DOR should be consistent with those for PFS as specified in Table 1 below.

PFS is defined as the time from randomization (or first dose date for non-randomized trials) to the first documented PD or death due to any cause, whichever comes first. Table 1 below contains the definitions of PFS events and the censoring rules for the primary analysis of PFS, where the efficacy reference start date is defined as the randomization date for randomized trials or the first dose date for non-randomized trials. Programmers would traverse these scenarios in sequence when determining PFS outcomes and dates. These definitions will be used to illustrate the ADaM modeling in Section 4.3.

Actual protocols may have slightly different calculations of DOR or PFS.

Table 1. PFS Event and Censoring Rules

| Step	Scenario	Outcome	Event or Censor Date
1	PD (no censoring criteria met)	Event	Date of PD
2	Death before first PD assessment (no censoring criteria met)	Event	Date of death
3	PD or death right after two or more consecutively missed tumor assessments and no other censoring criteria met	Censored	Date of last tumor assessment of CR, PR, or SD prior to the missed visits; if such tumor assessment does not exist, efficacy reference start date
4	Subsequent cancer-related therapy started before PD or death observed	Censored	Date of last tumor assessment of CR, PR, or SD on or prior to date of subsequent cancer-related therapy; if such tumor assessment does not exist, efficacy reference start date
5	No documented PD or death	Censored	Date of last tumor assessment of CR, PR, or SD; if such tumor assessment does not exist, efficacy reference start date

Besides the primary analysis of DOR or PFS, various sensitivity analyses using different definitions of event and censoring may be conducted depending on protocols. One example of the PFS sensitivity analysis is not to consider the subsequent cancer-related therapy as a censoring reason. Compared to the censoring criteria for the primary PFS as specified in Table 1 above, Step 4 will not be applicable for this specific sensitivity endpoint.

4. ADAM FOR EFFICACY ANALYSES BASED ON RECIST 1.1

4.1 OVERVIEW OF ADAM EFFICACY DATA SETS

According to SDTM IG 3.2 or above, tumor identification, tumor assessments, and tumor responses are mapped to the standard TU, TR, and RS domain, respectively. These SDTM domains together with other domains will then provide the input data to derive the RECIST 1.1 endpoints in Seagen's ADaM model, which include a total of five ADaM datasets: ADTR, ADRS, ADINTEV, ADEFFSUM, and ADTTE.

The diagram below shows the derivation order to create the five data sets after SDTM and ADSL are available. It also illustrates the data dependencies among the five data sets, where dotted arrows indicate conditional dependencies.
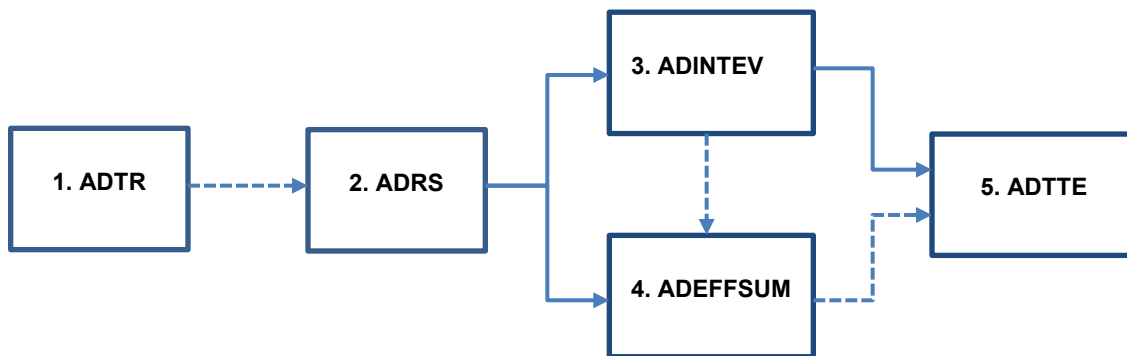
All five data sets use the ADaM basic data structure (BDS), where ADTR and ADRS are by-visit data, ADINTEV and ADEFFSUM are summary data at subject level, and ADTTE is TTE data.

- **ADTR**: Tumor Assessment Analysis Data contains data records with valid results, at both baseline and post-baseline time points, from the TR domain. Individual tumor assessment and SoD parameters are created in this data set.

- **ADRS**: Tumor Timepoint Response Analysis Data contains data records with valid results, at post-baseline time points only, from the RS domain if timepoint responses are provided by investigators or BICR; or it is derived from ADTR if timepoint responses are derived by the sponsor, and in this case, ADRS depends on ADTR.

- **ADINTEV**: Intermediate Event Analysis Data derives the intermediate PFS events including censoring events that contribute to the calculation of CBR, DOR, and PFS types of endpoints. ADINTEV depends on ADRS.

- **ADEFFSUM**: Efficacy Summary Analysis Data derives the categorical endpoints such as BOR, ORR, and CBR. ADEFFSUM depends on ADRS and, when the study includes CBR endpoints, it also depends on ADINTEV.

- **ADTTE**: Time-to-event Analysis Data derives DOR and PFS endpoints. ADTTE depends on ADINTEV and, when the study includes DOR endpoints, it also depends on ADEFFSUM.

4.2 ADTR AND ADRS DESIGN

This section uses some hypothetical data examples to illustrate the data structures for ADTR and ADRS. Efficacy endpoint derivations are similar between two data sources, INV vs. BICR. To simplify the concepts, all data examples and discussion focuses on INV data only.

Table 2 illustrates the data structure within ADTR where PARAMCD uniquely identifies the type of assessment, data source, and lesion ID where applicable. Many ADTR variables such as ABLFL, ADT, and ADY are excluded in Table 2 due to space constraints. Similarly, for the SoD parameters, both standard variables (BASE, CHG, and PCHG for baseline value, change from baseline, and percent change from baseline, respectively) and custom variables (NADIR, NCHG, and NPCHG for nadir value, change from nadir, and percent change from nadir, respectively) are also derived but not displayed in the table. Note that new lesions are assessed at the post-baseline visits only. Note that in many PARAMCDs, the third character (capital i) identifies these as investigator-based observations; if these were BICR-based, a B would be used instead.

Table 2. Example Data Records from ADTR

USUBJID	PARAMCD	PARAM	AVISIT	AVAL	AVALC
xyz-001	TSINW01	Tumor State of New Lesion NEW01 by INV	WEEK 6		EQUIVOCAL
xyz-001	TSINT01	Tumor State of Non-Target Lesion NT01 by INV	BASELINE		PRESENT

USUBJID	PARAMCD	PARAM	AVISIT	AVAL	AVALC
xyz-001	TSINT01	Tumor State of Non-Target Lesion NT01 by INV	WEEK 6		PRESENT
xyz-001	TSINT02	Tumor State of Non-Target Lesion NT02 by INV	BASELINE		PRESENT
xyz-001	TSINT02	Tumor State of Non-Target Lesion NT02 by INV	WEEK 6		ABSENT
xyz-001	TSINT03	Tumor State of Non-Target Lesion NT03 by INV	BASELINE		PRESENT
xyz-001	TSINT03	Tumor State of Non-Target Lesion NT03 by INV	WEEK 6		PRESENT
xyz-001	LDIT01	Longest Diameter of Non-nodal Target Lesion T01 by INV (mm)	BASELINE	30	
xyz-001	LDIT01	Longest Diameter of Non-nodal Target Lesion T01 by INV (mm)	WEEK 6	28	
xyz-001	LDIT02	Longest Diameter of Non-nodal Target Lesion T02 by INV (mm)	BASELINE	12	
xyz-001	LDIT02	Longest Diameter of Non-nodal Target Lesion T02 by INV (mm)	WEEK 6	12	
xyz-001	SODINV	Sum of Diameter by INV (mm)	BASELINE	42	
xyz-001	SODINV	Sum of Diameter by INV (mm)	WEEK 6	40	

Table 3 displays data records from ADRS. Four parameters are included in ADRS, target lesion response, non-target lesion response, new lesion progression, and overall response. Data source (INV vs BICR) is stored in a parameter qualifying variable named AEVAL. Endpoints such as BOR, ORR, and CBR are not created in ADRS yet; instead, they will be derived in ADEFFSUM. In ADRS, we also create a few flag variables to indicate the following:

1. Whether or not an assessment time point is on or before the start date of subsequent cancer-related therapy
2. Whether or not an assessment time point is on or before the first PD assessment
3. The first confirmed CR
4. The first confirmed PR
5. The first SD, CR, or PR assessment that meets the protocol specified-criteria (e.g., at least 5 weeks from the efficacy reference start date) for the BOR to be SD.

These flag variables are used for the data derivations in both ADINTEV and ADEFFSUM. They are not shown in Table 3.

Table 3. Example Data Records from ADRS

USUBJID	AEVAL	PARAMCD	PARAM	AVISIT	ADT	AVALC
xyz-001	INV	TRGRESP	Target Response	WEEK 6	27MAY2020	SD
xyz-001	INV	NTRGRESP	Non-target Response	WEEK 6	27MAY2020	NON-CR/NON-PD
xyz-001	INV	NEWLPROG	New Lesion Progression	WEEK 6	27MAY2020	N
xyz-001	INV	OVRLRESP	Overall Response	WEEK 6	27MAY2020	SD

4.3 ADINTEV AND ADTTE DESIGN

Derivation of PFS, CBR, and DOR types of endpoints are challenging and programming errors can occur due to the complex censoring rules. With the creation of the ADINTEV dataset, the complex PFS derivation is divided into intermediate components which eventually assemble the endpoint in ADTTE. The process with such an ADINTEV design is easier to document, understand, and program.

For the primary analysis of PFS with event and censoring defined in Section 3.2, we would create the intermediate events listed in Table 4.1.

Table 4.1 Intermediate Events for Primary Analysis of PFS

AEVLINT	PARAMCD	PARAM
OVERALL	EFFRFS	Efficacy Reference Start

AEVLINT	PARAMCD	PARAM
OVERALL	NCTX	New Cancer Therapy
ON OR BEFORE NEW CANCER THERAPY	DEATH	Death
ON OR BEFORE NEW CANCER THERAPY	PD	Progressive Disease
ON OR BEFORE NEW CANCER THERAPY	LETAPD	Last Evaluable Tumor Assessment Before PD
ON OR BEFORE NEW CANCER THERAPY	M2TAPD	Missed >= 2 Consecutive Tumor Assessments Right Before PD or Death

And for the sensitivity analysis of PFS that ignores the use of subsequent cancer-related therapy, we would add additional intermediate events listed in Table 4.2. Compared to the primary PFS, the same PARAMCD and PARAM are used; but it no longer requires ON OR BEFORE NEW CANCER THERAPY, instead, the analysis interval would be the entire study, i.e., OVERALL.

Table 4.2 Intermediate Events Additionally Needed for Sensitivity Analysis of PFS

AEVLINT	PARAMCD	PARAM
OVERALL	DEATH	Death
OVERALL	PD	Progressive Disease
OVERALL	LETAPD	Last Evaluable Tumor Assessment Before PD
OVERALL	M2TAPD	Missed >= 2 Consecutive Tumor Assessments Right Before PD or Death

Table 5 shows a few example data records from ADINTEV supporting the primary PFS derivation. When an intermediate event occurs, AVALC is assigned to Y; otherwise, AVALC is assigned to N. This approach makes data more explicit to distinguish between two scenarios: when a given event does not occur vs. when a given event is mistakenly not programmed.

Table 5. Example Data Records from ADINTEV

USUBJID	AEVAL	AEVLINT	PARAMCD	AVALC	ADT
xyz-002	INV	OVERALL	EFFRFS	Y	03JAN2020
xyz-002	INV	OVERALL	NCTX	Y	17MAY2021
xyz-002	INV	ON OR BEFORE NEW CANCER THERAPY	DEATH	N	
xyz-002	INV	ON OR BEFORE NEW CANCER THERAPY	PD	Y	10MAY2021
xyz-002	INV	ON OR BEFORE NEW CANCER THERAPY	LETAPD	Y	21DEC2020
xyz-002	INV	ON OR BEFORE NEW CANCER THERAPY	M2TAPD	Y	

Using these data examples from Table 5 above, derivation of PFS in ADTTE concludes that subject xyz-002 is censored at the last evaluable tumor assessment on or before the PD, which is 21DEC2020. The data derivation logic is as follows: if (PD = Y or DEATH = Y) and M2TAPD ^= Y then PD or death event occurred; else if (PD = Y or DEATH = Y) and M2TAPD = Y then we censor due to missing visits; else if NCTX = Y then we censor due to new cancer therapy; else we censor due to no PD or death. The data record in ADTTE for this example would look like:

USUBJID	AEVAL	PARAMCD	STARTDT	ADT	AVAL	AVALU	CNSR	EVNTDESC
xyz-002	INV	PFS	03JAN2020	21DEC2020	11.63	MONTHS	1	EVENT RIGHT AFTER >= 2 CONSECUTIVELY MISSED TUMOR ASSESSMENTS

4.4 ADEFFSUM DESIGN

ADEFFSUM is a summary type of data set (like ADINTEV) that contains categorical endpoints, BOR, ORR, CBR, etc. Table 6 displays a few data examples in ADEFFSUM. Note that ADT and ADY are derived for ORR only where ADY is the time to the objective response in days.

Table 6. Example Data Records from ADEFFSUM

USUBJID	AEVAL	AEVLINT	PARAMCD	AVALC	ADT	ADY
xyz-003	INV	ON OR BEFORE NEW CANCER THERAPY	BOR	PR		
xyz-003	INV	ON OR BEFORE NEW CANCER THERAPY	ORR	Y	05APR2019	167
xyz-003	INV	ON OR BEFORE NEW CANCER THERAPY	CBR	Y		

Some may choose to put these endpoints directly into ADRS, but we have found separating the derivations of timepoint responses and categorical endpoints into two datasets, ADRS and ADEFFSUM respectively, have the following benefits.

1. While timepoint response data require a more comprehensive BDS structure with a longer list of variables as detailed in Section 4.2, the summary data set is a simpler BDS design with only a handful of domain-specific variables needed, such as AEVLINT, AEVAL, PARAM, PARAMCD, AVALC (and / or AVAL), ADT, and ADY. Separating ADRS and ADEFFSUM simplifies both data sets and makes them easier for reviewers to understand.

2. Endpoint data records depend on timepoint response records. Having record dependencies within a same data set will complicate the metadata and programming. Separating ADRS and ADEFFSUM provides step-by-step traceability in the metadata.

3. It allows endpoint derivations to use ADINTEV as the input when ADEFFSUM includes the CBR parameters. As explained in Section 4.3, ADINTEV provides efficient intermediate components to derive time to disease progression, which is part of the CBR definition. Since ADINTEV also depends on ADRS, it would be impossible for CBR processing to take advantage of the ADINTEV design if CBR is derived within ADRS.

Overall, separating visit-based data (ADRS) from summary-based data (ADEFFSUM) improves reviewability, and programming efficiency and quality.

4.5 USE OF PARAMETER QUALIFYING VARIABLES

As seen in the data examples from Section 4.2, 4.3, and 4.4, parameter qualifying variables are introduced into the BDS structures, where feasible. ADRS, ADINTEV, ADTTE, and ADEFFSUM are structured by parameter qualifying variables (AEVAL and AEVLINT where applicable) in addition to the PARAM and PARAMCD variables. The parameter qualifying variables provide natural layers within the data and simplify the design of analysis parameters. For example, in the ADINTEV data set, the variable AEVLINT specifies the analysis intervals (OVERALL vs. ON OR BEFORE NEW CANCER THERAPY) and the variable AEVAL identifies the data source (INV vs. BICR). The qualifying information is not part of the analysis parameters.

Use of parameter qualifying variables impacts the ADaM value-level metadata (VLM). Instead of displaying one-level VLM using PARAMCD only, multiple-level VLM combining PARAMCD and applicable qualifying variable(s) may be more helpful and provide better clarity. For example, to define AVALC in the ADINTEV dataset, it is helpful to describe the AVALC derivation algorithms by each combination of PARAMCD and AEVLINT. Note that as data derivations including controlled terminology (CT) are similar between INV and BICR data, most of the VLM does not need to involve the qualifying variable AEVAL.

When generating tables, listings, and figures (TLFs) from the ADaM data sets, the programmer must select an appropriate PARAMCD value as well as a value of any parameter qualifying variable used in the ADaM data set. For example, an analysis based on ADEFFSUM.AVALC would subset based on PARAMCD = BOR, AEVLINT = ON OR BEFORE NEW CANCER THERAPY, and AEVAL = INV.

5. ADAM PROGRAMMING STANDARDIZATION

Based on the standard ADaM structures, we are able to standardize common and complex data derivations as departmental macros. Use of macros further improves programming efficiency and quality.

5.1 DERIVATION OF THE TIMEPOINT RESPONSE

Although timepoint responses are usually summarized and reported by investigators or BICR, sponsors may also need to derive them from the individual tumor assessments to: 1) fulfil an FDA information request (IR); and 2) conduct data quality checks.

Programming of timepoint responses is extremely complex per RECIST 1.1. However, with standardized data collection and SDTM and ADaM structures, it is a good candidate for departmental macros. In Seagen's implementation, a series of macros are developed to work with the standard ADaM models and to support the programming process:

- A macro that creates the complete ADTR structure from the SDTM TR domain including the derivation of the SoD records based on the measurements of individual target lesions.

- A modular macro that derives BASE, CHG, PCHG, NADIR, NCHG, and NPCHG on the SoD parameters after they are created in ADTR.

- A macro that derives all the parameters defined in ADRS (target lesion response, non-target lesion response, new lesion progression, and overall response) based on ADTR.

- A modular macro that derives the two flag variables in ADRS, the first confirmed CR and the first confirmed PR, after the overall response parameter is created in ADRS.

5.2 DERIVATION OF PROGRESSION-FREE SURVIVAL

The derivation of PFS or DOR endpoints in ADTTE becomes standard especially when consistent definitions are used across studies. We are able to use a departmental macro for the programming of PFS or DOR using the components derived in the ADINTEV data set as the input. The same macro can also be used for deriving the SD duration which is part of the CBR algorithm.

6. CONCLUSION

The ADaM structures described in this paper enable company-wide programming standardization of the complex derivations of timepoint tumor response based on RECIST 1.1 and PFS type of endpoints including various versions of PFS sensitivity analysis endpoints. With the creation of departmental data derivation macros, programming efficiency and quality can be greatly improved.

These structures can be used to handle various protocol designs (e.g., one-treatment-period design vs. treatment crossover design) and both data sources (i.e., data reported by investigators and by BICR).

These structures can also be expanded for iRECIST, which was developed by the RECIST working group for use in cancer immunotherapy trials.

Having implemented these structures in pivotal trials supporting multiple submissions, Seagen found that this easily accessible and reviewable analysis and data package may have been of benefit to Agency reviewers, as such filings were reviewed smoothly with fewer information requests than seen on prior submissions.

7. RECOMMENDED READING

- *RECIST Guideline (Version 1.1)*

- *CDISC SDTM Implementation Guide V3.2 (or above)*

- *CDISC ADaM Implementation Guide V1.1 (or above)*

8. ACKNOWLEDGMENTS

Authors would like to thank the Seagen ADaM Standards Committee, Tukysa biometrics team, and Statistical Programming Infrastructure (SPI) team. Their innovative designs and successful implementations provided great ideas and examples that are summarized in this paper.

9. CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Ellen Lin
Seagen Inc.
ellin@seagen.com

Matt Ness
Seagen Inc.
mness@seagen.com

Any brand and product names are trademarks of their respective companies.