# Programming Challenges in Developing PRO Analysis Datasets Under FDA's New Submission Guidance

Weiwei Guo, Statistical Programmer, Merck & Co., Inc., Rahway, NJ, USA
David Huang, Statistician, Merck & Co., Inc., Upper Gwynedd, PA, USA

## ABSTRACT

Patient-Reported Outcomes (PRO) data are measurements of a patient's health status, symptoms, functional abilities, health-related quality of life, and treatment satisfaction that are captured directly from the patient, without clinician interpretation. Because PRO collection depends on active patient participation, missing data is common, posing substantial challenges for statistical programming and downstream analysis. In November 2023, the U.S. Food and Drug Administration (FDA) issued "Submitting Patient-Reported Outcome (PRO) Data in Cancer Clinical Trials," which outlines expectations for CDISC SDTM and ADaM datasets and recommends specific tables, listings, and figures (TLFs) for oncology submissions. In September 2025, our study team received an FDA information request to provide PRO TLFs in accordance with this guidance. This paper describes programming challenges and practical solutions for constructing analysis-ready PRO datasets, focusing on handling missing item-level responses and derived summary scores, and on deriving key visit-level flags to support the TLFs requested in the FDA's new guidance.

## BACKGROUND

In November 2023, the U.S. Food and Drug Administration (FDA) issued guidance titled "Submitting Patient-Reported Outcome Data in Cancer Clinical Trials." The guideline provides specifications for the submission of Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM) datasets and specifications for recommended tables and figures.

In September 2025, our study team received an FDA information request to submit PRO tables, listings, and figures (TLFs) in the format specified by this guidance. The request included patient disposition, PRO data completeness, and distribution tables and figures for evaluating safety, tolerability, and clinical benefit. To produce these TLFs, we first need to build analysis-ready PRO datasets that meet the requirements for the requested tables and figures. This paper

describes programming challenges and practical solutions for creating analysis-ready PRO datasets. We focus on handling missing item-level responses and derived summary scores, and on deriving key visit-level flags needed to support the TLFs specified in the FDA's new guidance.

## TLFs REQUIRED BY FDA GUIDANCE

The new FDA guidance specifies TLFs covering three key areas: patient disposition, PRO data completeness, and PRO score distributions. Each of these areas requires tables and figures to support the assessment of both clinical benefit and safety/tolerability.

Example Table A4 presents the FDA mock-up for a patient disposition table used to evaluate clinical benefit. In this table, the randomized population serves as the denominator. For each visit (e.g., Baseline, Cycle 2 Day 1, Cycle 3 Day 1), the total number of patients is fixed to the number randomized, and all categories under "PRO Expected" and "PRO Not Expected" must sum to the randomized count for the corresponding arm and visit.

"PRO Expected" typically includes patients who remain on therapy as well as those who discontinue for reasons other than death, regardless of whether they actually complete the PRO assessment. "PRO Not Expected" includes patients for whom PRO data are not anticipated at that timepoint—for example, due to death or other reasons such as lack of an appropriate language translation. In the ideal baseline scenario illustrated in mock-up A4, where all randomized patients receive treatment, the PRO is expected for everyone; therefore, the entire randomized cohort appears in the "Patients on Therapy" column. At subsequent visits, patients who discontinue because of disease progression, adverse events, other reasons, or who die are assigned to the corresponding categories, ensuring that the sum of all categories equals the randomized population for each visit and treatment arm.

# Example Tables and Figures in new FDA Guidance

PROEX1FL

PNEX1CAT

ONTRTFL

EOTSTCAT

DTHRFL

**Table A4. Patient Disposition when Evaluating Clinical Benefit (Denominator = Randomized Population)[1]**

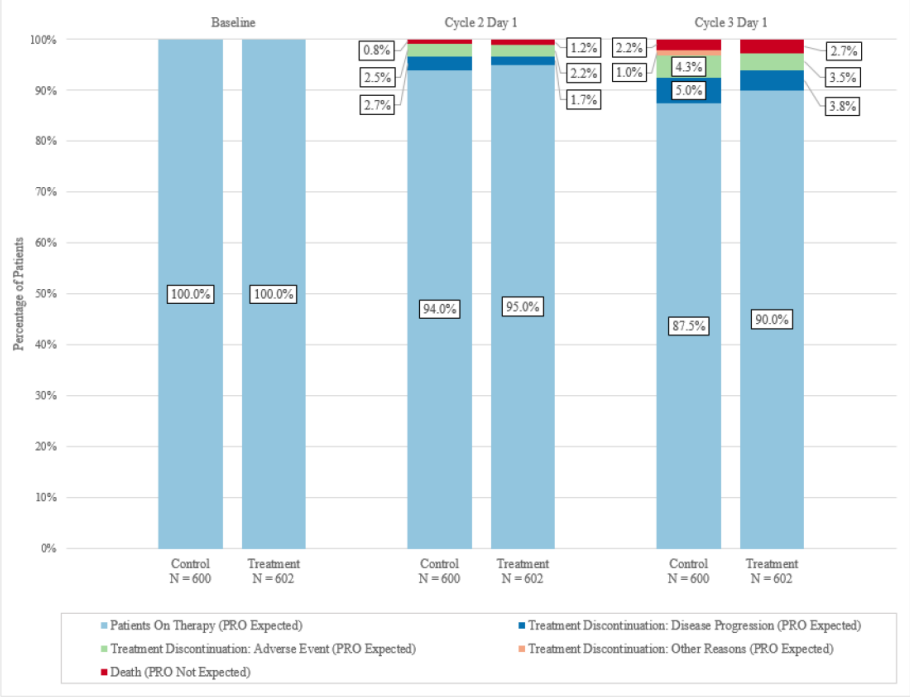
| Analysis Visit | Treatment Arm | Randomized Patients (N) | PRO Expected[2] | | | | PRO Not Expected | |
|---|---|---|---|---|---|---|---|---|
| | | | Patients On Therapy, n (%) | Treatment Discontinuation: Disease Progression, n (%) | Treatment Discontinuation: Adverse Event (AE), n (%) | Treatment Discontinuation: Other Reasons, n (%) | Death, n (%) | Other,[3] n (%) |
| Baseline | Control | 600 | 600 (100.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) |
| | Treatment | 602 | 602 (100.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) |
| Cycle 2 Day 1 | Control | 600 | 564 (94.0%) | 16 (2.7%) | 15 (2.5%) | 0 (0.0%) | 5 (0.8%) | 0 (0.0%) |
| | Treatment | 602 | 572 (95.0%) | 10 (1.7%) | 13 (2.2%) | 0 (0.0%) | 7 (1.2%) | 0 (0.0%) |
| Cycle 3 Day 1 | Control | 600 | 525 (87.5%) | 30 (5.0%) | 26 (4.3%) | 6 (1.0%) | 13 (2.2%) | 0 (0.0%) |
| | Treatment | 602 | 542 (90.0%) | 23 (3.8%) | 21 (3.5%) | 0 (0.0%) | 16 (2.7%) | 0 (0.0%) |

Randomized

All columns add-up equal to randomized

[1] Denominator used to calculate percentages is the number of randomized patients.
[2] When PRO data are used to evaluate clinical benefit, the PRO measure is generally expected to be completed for patients who both did not discontinue treatment and who discontinued treatment for reasons other than death. Columns under PRO Expected includes patients who both completed and did not complete the PRO measure (e.g., the patient did not attend an onsite visit, the patient did not complete the PRO measure at the attended onsite visit or at a prespecified offsite assessment timepoint, the patient partially completed the PRO measure resulting in incalculable summary scores).
[3] The Other column groups patients not expected to complete the PRO measure at a designated assessment timepoint for reasons other than patient death (e.g., the translation of the PRO measure is not available in the patient's language).

**Figure A1. Patient Disposition when Evaluating Clinical Benefit (Denominator = Randomized Population)**



#### 5.3.2 *Patient Disposition when Informing Safety and Tolerability*

**Table A5. Patient Disposition when Informing the Evaluation of Safety and Tolerability (Denominator = Safety Population)[4]**

| Analysis Visit | Treatment Arm | Randomized Population (N) | Safety Population (N) | PRO Expected,[5] n (%) | PRO Not Expected | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | Death, n (%) | Treatment Discontinuation: Disease Progression, n (%) | Treatment Discontinuation: Adverse Event, n (%) | Treatment Discontinuation: Other Reasons, n (%) | Other,[6] n (%) |
| Baseline | Control | 600 | 600 | 600 (100.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) |
| | Treatment | 602 | 602 | 602 (100.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) |
| Cycle 2 Day 1 | Control | 600 | 600 | 564 (94.0%) | 5 (0.8%) | 16 (2.7%) | 15 (2.5%) | 0 (0.0%) | 0 (0.0%) |
| | Treatment | 602 | 602 | 572 (95.0%) | 7 (1.2%) | 10 (1.7%) | 13 (2.2%) | 0 (0.0%) | 0 (0.0%) |
| Cycle 3 Day 1 | Control | 600 | 600 | 525 (87.5%) | 13 (2.2%) | 30 (5.0%) | 26 (4.3%) | 6 (1.0%) | 0 (0.0%) |
| | Treatment | 602 | 602 | 542 (90.0%) | 16 (2.7%) | 23 (3.8%) | 21 (3.5%) | 0 (0.0%) | 0 (0.0%) |

[4] Denominator used to calculate percentages is the number of patients in the safety population.
[5] When PRO data are used to inform the evaluation of safety and tolerability, the PRO measure may not be expected to be completed after a patient discontinues from treatment as shown in the table. Thus, the PRO Expected column excludes patients who discontinued from treatment. The PRO Expected column is determined where PRO Expected Flag (PROEXPFL) equals 'Y' and includes patients who both completed and did not complete the PRO measure (e.g., the patient did not attend an onsite visit, the patient did not complete the PRO measure at the attended onsite visit or at a prespecified offsite assessment timepoint, the patient partially completed the PRO measure resulting in incalculable summary scores).
[6] The Other column groups patients who were not expected to complete the PRO measure at a designated assessment timepoint for reasons other than treatment discontinuation or patient death (e.g., the translation of the PRO measure is not available in the patient's language).

## THE ACTUAL COLLECTED SDTM QS DATA

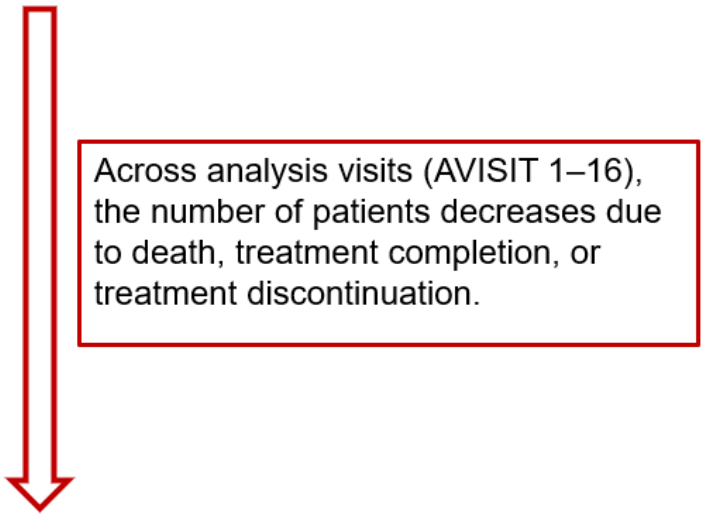
Table 1 presents the PROC FREQ output for a single PRO item score based on the collected SDTM QS data. In this example, the number of patients with available data decreases across AVISIT 1–16 as a result of death, treatment completion, or treatment discontinuation.

In contrast, the FDA mock-up requires that the TLFs display all patients in the randomized or safety population at every visit. Thus, the analysis denominator must remain constant across visits, even when PRO data are not collected or are missing.

To satisfy these TLF requirements, patients with missing PRO data at a given visit must be represented in the ADaM dataset through the creation of phantom records. These derived records ensure that every patient in the relevant analysis population is included at each visit, allowing the TLFs to fully align with the structure specified in the FDA guidance.

**Table 1: The Actual SDTM QS Data**

| AVISIT | PATIENT |
|---|---|
| 1 | 600 |
| 2 | 586 |
| 3 | 561 |
| 4 | 478 |
| 5 | 430 |
| 6 | 387 |
| 7 | 311 |
| 8 | 233 |
| 9 | 203 |
| 10 | 178 |
| 11 | 147 |
| 12 | 121 |
| 13 | 100 |
| 14 | 88 |
| 15 | 78 |
| 16 | 58 |




## HOW TO DERIVE PHANTOM RECORDS IN ANALYSIS DATASET FOR MISSING DATA?

Per FDA new guidance "*Missing PRO data should be represented within the QS dataset with the reason for missingness captured under QS. QSREASND. The QS dataset should include one record per item per PRO measure per patient per assessment timepoint, regardless of whether an item response is missing*". For the missing item score does not exist in the SDTM QS domain, we must derive the **PHANTOM** records in ADaM dataset.

For the programming implementation, phantom records are derived at each assessment timepoint (pre-defined visit window per SAP) using a visit-frame approach. Each PRO item or summary score is represented by a unique PARAM/PARAMCD. Phantom records are then generated for every visit–instrument–item combination within this visit frame, ensuring that all expected assessments are represented in the ADaM dataset, regardless of whether PRO data were actually collected.

## HOW TO DERIVE THE KEY FLAGS IN ADaM PER EACH VISIT

As shown in Example Table A4, each column is represented by a single variable from the ADaM dataset. These variables are visit-based, so their values may change over time for a given patient. The following variables are added in the ADaM dataset to support the TLF analysis needs:

- PROEX1FL: PRO expectation flag for evaluating clinical benefit.
- PROEX2FL: PRO expectation flag for safety and tolerability
- PNEX1CAT: Category for PRO not expected (clinical benefit)
- PNEX2CAT: Category for PRO not expected (safety and tolerability)
- PROSCMFL: PRO summary score completed flag, per instrument-specific rules.
- DTHRFL: Death flag by visit
- ANLxxFL: PRO Analysis flag
- ONTRTFL: On-treatment record flag
- AREASND: Analysis reason not performed (carried forward or derived from QSREASND)
- EOTSTCAT: Treatment status category
- DTYPE: Populate “Phantom” for phantom records

All of these variables are visit-based. For example, the PRO ADaM dataset includes a death flag, DTHRFL, which differs from the death flag in ADSL. In ADSL, the death flag is subject level: for each patient it is either Y or N and does not vary by visit. In PRO ADaM, DTHRFL is visit-level. As shown in Example Table 4, for the first two visits the patient is alive, so DTHRFL = 'N'. The patient dies at the second visit, and from visit 3 onward DTHRFL = 'Y'.

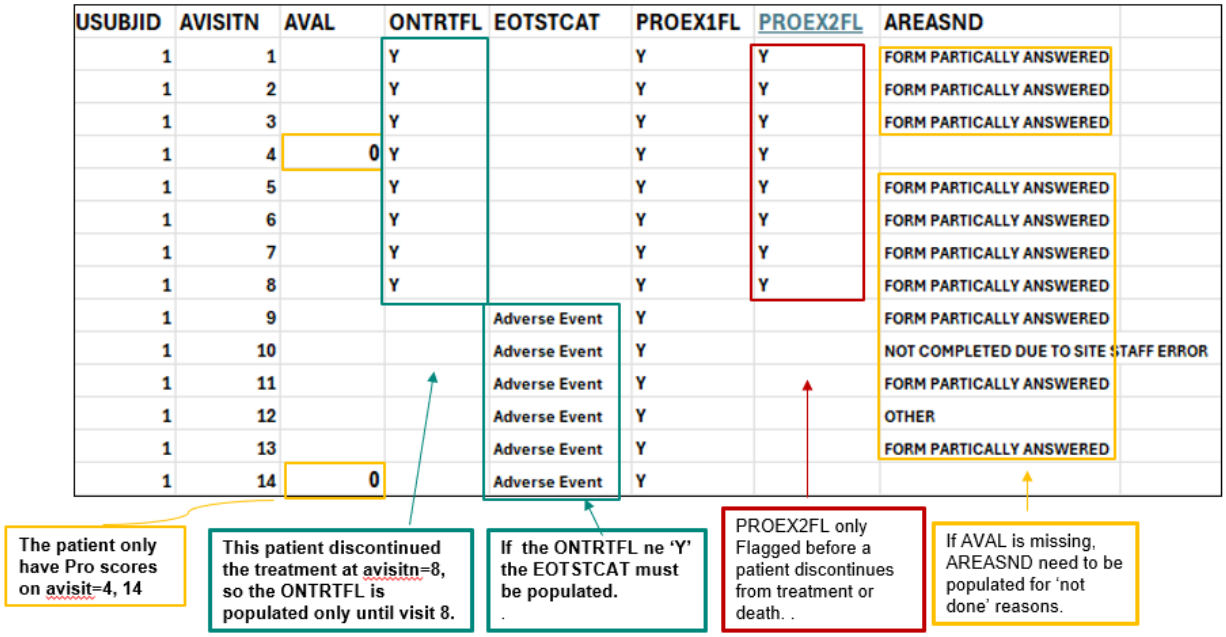
These variables are also interrelated within each visit window. Using Example Table A4 (the disposition table for evaluating clinical benefit), the relationships can be illustrated as follows: for each visit, the total patient count must equal the number randomized. The “PRO expected” / “PRO not expected” columns and their subcategories must sum exactly to this total for that visit. For example, at baseline all patients are on treatment, so the on-treatment count should equal the randomized number. At later visits, some patients may have discontinued treatment or died;

those patients must be classified into the appropriate “discontinuation” or “death” categories. Consequently, each variable defining these categories and subcategories must be clearly defined and mutually exclusive so that row totals always match the randomized population.

Table 3 provides an example of the ADaM structure for a single patient and a single PRO score. For this patient and this PRO item, PRO data are only available at visits 4 and 14, as indicated by non-missing values in AVAL. For all visits with missing AVAL, the AREASND variable (the last column) must be populated with the appropriate “not done” reason. For the columns highlighted in green, if ONTRTFL is not "Y", EOTSTCAT must be populated with the relevant discontinuation reason. In addition, PROEX1FL and PROEX2FL are defined differently for patients who are randomized but not treated, and PROEX2FL flags only those visits occurring before treatment discontinuation or death.

Table 4 shows another example of the ADaM structure for a single patient. For this patient, visits 2, 3, 4, and 6 are not scheduled per the study SAP. The patient has a PRO measure only on the first visit and dies on the second visit. Phantom records must therefore be derived for all visits after the patient’s death. As highlighted in the green columns, whenever ONTRTFL is not "Y", either EOTSTCAT or the death flag must be populated.

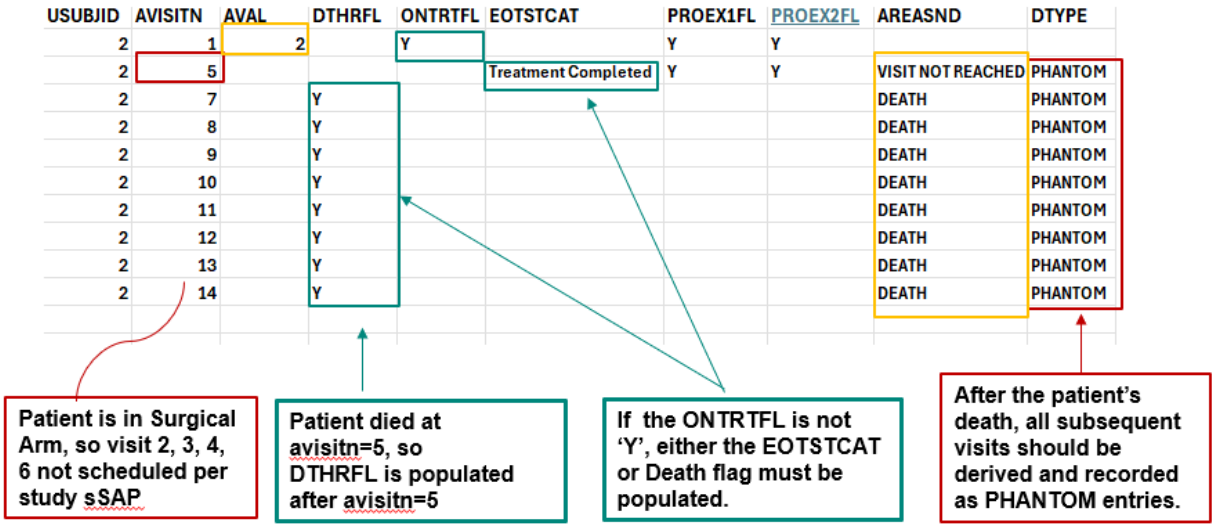
**Table 3: Example of ADaM structure for key flags by each visit window**



| USUBJID | AVISITN | AVAL | ONTRTFL | EOTSTCAT | PROEX1FL | PROEX2FL | AREASND |
|---|---|---|---|---|---|---|---|
| 1 | 1 | | Y | | Y | Y | FORM PARTICALLY ANSWERED |
| 1 | 2 | | Y | | Y | Y | FORM PARTICALLY ANSWERED |
| 1 | 3 | | Y | | Y | Y | FORM PARTICALLY ANSWERED |
| 1 | 4 | 0 | Y | | Y | Y | |
| 1 | 5 | | Y | | Y | Y | FORM PARTICALLY ANSWERED |
| 1 | 6 | | Y | | Y | Y | FORM PARTICALLY ANSWERED |
| 1 | 7 | | Y | | Y | Y | FORM PARTICALLY ANSWERED |
| 1 | 8 | | Y | | Y | Y | FORM PARTICALLY ANSWERED |
| 1 | 9 | | | Adverse Event | Y | | FORM PARTICALLY ANSWERED |
| 1 | 10 | | | Adverse Event | Y | | NOT COMPLETED DUE TO SITE STAFF ERROR |
| 1 | 11 | | | Adverse Event | Y | | FORM PARTICALLY ANSWERED |
| 1 | 12 | | | Adverse Event | Y | | OTHER |
| 1 | 13 | | | Adverse Event | Y | | FORM PARTICALLY ANSWERED |
| 1 | 14 | 0 | | Adverse Event | Y | | |



**Table 4: Example of ADaM structure for key flags by each visit window**



| USUBJID | AVISITN | AVAL | DTHRFL | ONTRTFL | EOTSTCAT | PROEX1FL | PROEX2FL | AREASND | DTYPE |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 1 | 2 | | Y | | Y | Y | | |
| 2 | 5 | | | | Treatment Completed | Y | Y | VISIT NOT REACHED | PHANTOM |
| 2 | 7 | | Y | | | | | DEATH | PHANTOM |
| 2 | 8 | | Y | | | | | DEATH | PHANTOM |
| 2 | 9 | | Y | | | | | DEATH | PHANTOM |
| 2 | 10 | | Y | | | | | DEATH | PHANTOM |
| 2 | 11 | | Y | | | | | DEATH | PHANTOM |
| 2 | 12 | | Y | | | | | DEATH | PHANTOM |
| 2 | 13 | | Y | | | | | DEATH | PHANTOM |
| 2 | 14 | | Y | | | | | DEATH | PHANTOM |

## CONCLUSIONS

In summary, to support the TLFs required by the new FDA guidance, the ADaM implementation must meet several key requirements:

1. **Population handling:** The dataset must simultaneously support analyses for both the randomized and safety populations, enabling consistent evaluation of clinical benefit and safety/tolerability.
2. **Phantom record derivation:** Phantom records must be derived for missing PRO scores, with at least one record per item, per PRO measure, per patient, per assessment timepoint.
3. **Addition of visit-based variables:** New visit-based variables must be created to define analysis categories and subcategories. These variables are derived using visit windows and must be clearly specified and mutually exclusive so that, for each visit, the row totals align exactly with the randomized population and safety population.

As a consequence, the ADPRO dataset can become very large once phantom records are introduced at the item/PRO-measure/patient/timepoint level. For example, a study with 700 randomized patients, 14 visits, and 100 parameters can result in nearly one million records. Within this expanded structure, it is critical that each derived variable be precisely defined and mutually exclusive across all visits to maintain internal consistency and interpretability of the analyses.

## REFERENCES

**"Submitting Patient-Reported Outcome Data in Cancer Clinical Trials**", Guidance for Industry, Technical Specifications Document.

https://www.fda.gov/regulatory-information/search-fda-guidance-documents/submitting-patient-reported-outcome-data-cancer-clinical-trials

## ACKNOWLEDGMENTS

The authors would like to thank Merck management team for their review of this paper and great support for this urgent FDA PRO IR request.

## CONTACT INFORMATION

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

Name: Weiwei Guo
Enterprise:  Merck & Co., Inc.
Address: 126 E. Lincoln Avenue, Rahway, NJ, 07065
Email: weiwei.guo@merck.com

Name: David Huang
Enterprise: Merck & Co., Inc.
Address: 351 N Sumneytown Pike, North Wales, PA, 19454
Email: david.huang@merck.com