Use of Diaries to Capture Patient-Reported Outcomes in FDA New Drug Approvals (2019-2023)

Sarah Sauchelli,[1,*] Emily Chicklis,[2] Michelle K Carty,[2] Ari Gnanasakthy[3]

[1] RTI Health Solutions, Manchester, United Kingdom; [2] IQVIA, Durham, NC, United States; [3] RTI Health Solutions, Research Triangle Park, NC, United States

PCR94

INTRODUCTION

- Endpoints derived from diaries have been recommended by the Food and Drug Administration (FDA) for the development of new products for certain diseases (e.g., preventative treatment of migraines[1]).
- To date, FDA guidance on the development and implementation of diaries to support patient-reported outcome (PRO)–based labeling is limited to the broader patient-focused drug development guidance documents.[2-4]
- The objectives of this review were two-fold:
  1. Examine use of diaries in pivotal clinical trials supporting treatment efficacy in new products approved by the FDA
  2. Inspect FDA review and approval of diary-based PRO data for inclusion in the product label

METHODS

- New molecular entities approved by the FDA between 2019 and 2023 were identified.
- The FDA labels and review documents were analyzed to identify and extract relevant information pertaining to the use of diaries to assess PRO endpoints in pivotal trials.
- Reviewer comments pertaining to the development and implementation of the diaries, as well as analysis and interpretation of diary-derived PRO data, were organized into thematic categories.

RESULTS

- Diaries were used to collect PRO data in support of an endpoint for 15% (n = 37) of 241 drugs approved by the FDA between 2019 and 2023 (Table 1).
  - Diary-based data in pivotal trials were most frequent for drugs to treat diseases of the nervous system (n = 15/37, 41%; 8 unique diseases).
- Overall approval of diary-assessed data for labeling purposes was high (n = 30/37, 81% of drugs), although there was variability across therapeutic areas (Table 1).
- Where data collection mode was reported (n = 24), electronic diaries were more frequently used than paper-based diaries (21 [88%] vs. 3 [13%] instances, respectively).
- Diaries were often included in pivotal trial(s) to capture 1 or more PROs assessed as primary and/or key secondary endpoints (Table 2).
  - Endpoint placement appeared to have a negligible impact on approval of diary-based data in the label.
- FDA feedback on diary-based data included consideration of the suitability of the questionnaire, study design, data quality, analysis of data, and interpretability of results (Table 3).
  - Feedback on diary-based data generally align with the types of critical comments identified for all PRO data.[5]
  - Handling of missing diary data (including predefined missing data rules) and the inclusion of multiple endpoints assessed via a diary are important components of the FDA's evaluation of diary-based data.

Table 1. Diary Use to Collect PRO Data in Pivotal Trials to Demonstrate Treatment Effect Across Therapeutic Specialties

Therapeutic area (ICD-10)	Drugs approved N	Diary used to collect PRO data n (% of drugs)	Label based on diary data n (% of instances of diary use)	Example diary-based concepts approved for labeling[a]
Certain infectious and parasitic diseases	20	—	—	—
Codes for special purposes	1	1 (100.0)	1 (100.0)	***"COVID-19 signs and symptoms [...]"***
Congenital malformations, deformations, and chromosomal abnormalities	5	1 (20.0)	1 (100.0)	*"**Diarrhea** occurred in 77% of 74 pediatric patients"*
Diseases of the blood and blood-forming organs	16	4 (25.0)	3 (75.0)	*"...produced statistically significant reductions in the rate of **HAE attacks** compared to placebo"*
Diseases of the circulatory system	4	—	—	—
Diseases of the digestive system	5	3 (60.0)	3 (100.0)	*"...improvements from baseline in average weekly **CSBMs** and **abdominal pain** were observed by week 1"*
Diseases of the eye and adnexa	6	—	—	—
Diseases of the genitourinary system	5	3 (60.0)	3 (100.0)	*"...statistically significant reduction [...] moderate to severe **vasomotor symptoms**"*
Diseases of the musculoskeletal system and connective tissue	6	—	—	—
Diseases of the nervous system	37	15 (40.5)	13 (86.7)	*"...the percentage of patients achieving **headache pain freedom** and **MBS freedom**..."*
Diseases of the respiratory system	3	—	—	—
Diseases of the skin and subcutaneous tissue	14	3 (21.4)	3 (100.0)	*"A greater proportion of subjects [...] achieved PSSD symptom score of 0 (absence of **itch, pain, burning, stinging, and skin tightness**)"*
Endocrine, nutritional, and metabolic diseases	28	3 (10.7)	2 (66.7)	*"**Hunger** scores [...] improved when IMCIVREE was reinitiated"*
Factors influencing health status and contact with health services	3	—	—	—
Injury, poisoning, and certain other consequences of external causes	1	—	—	—
Mental and behavioral disorders	9	—	—	—
Neoplasms	71	4 (5.6)	1 (25.0)	*"The MFS result was supported by a delay in time to pain progression"*
Others	1	—	—	—
Pregnancy, childbirth, and the puerperium	1	—	—	—
Symptoms, signs, and abnormal clinical and laboratory findings	5	—	—	—
Total drugs	**241**	**37 (15.0)**	**30 (81.1)**	

COVID-19 = coronavirus disease 2019; CSBM = complete spontaneous bowel movement; HAE = hereditary angioedema; ICD-10 = *International Classification of Diseases, Tenth Revision*; MBS = most bothersome symptom; MFS = metastasis-free survival; PSSD = Psoriasis Symptoms and Signs Diary.

[a] PRO concepts in bold.

Table 2. Inclusion of Diary-Assessed PROs in Labels Across Endpoint Type

Endpoint placement	Drugs with diary-based PRO data (N = 37)	Drugs with diary-based PRO concepts in labels[a] (N = 30)
Primary, n (%)	21 (56.8)	21 (70.0)
Key secondary, n (%)	33 (89.1)	28 (93.3)
Other endpoint,[b] n (%)	9 (24.3)	7 (23.3)

Note: More than 1 type of diary-assessed endpoint may have been used in a clinical trial and/or included in a drug label. For 1 product, the diary was used to assess tolerability of the drug rather than clinical efficacy, so it has been excluded.

[a] Counts are based on at least 1 endpoint of each category having been included in the product label.

[b] Includes endpoints labeled as "exploratory endpoints."

Table 3. FDA Considerations When Evaluating Diary-Assessed PRO Data

Theme	Types of FDA feedback	Description
Instrument	Fitness-for-purpose	Whether the diary or the PRO measure administered as a diary is fit-for-purpose
	Comprehensiveness	Whether the diary captures all PRO concepts of importance to the target population
	Device validation and implementation	Validation and implementation of the device to administer the diary measure
Study design	Overlap in endpoints	Impact of conceptual overlap in diary-assessed endpoints, in terms of patients' ability to distinguish between concepts, potential recall bias, and/or interpretability of results
	Endpoint appropriateness	Appropriateness of endpoint, including reference to previous uses of the endpoint to support product approvals in the disease and/or for the drug
	Timing of administration	Whether the schedule of assessment is appropriate
	Responder bias	Potential biases associated with diaries being patient reported
	Single-arm study	Potential impact or mitigation of risk of bias in data collected from a single-arm study
	Recall bias	Handling of potential recall bias when diary data are collected retrospectively
Data quality	Unblinding	Handling of any potential unblinding and/or concerns on the impact of treatment unblinding
	Completion compliance	Concerns around completion compliance (including how it has been defined)
	Error(s) in data entry	Inconsistencies between the data entered in an electronic diary and data entered in the case report form, as well as errors identified in diary data entry
Analysis	Adjustment for multiplicity	Methods used to adjust analyses for multiplicity, if any have been applied, including the sponsor's decisions on which diary-based endpoints to include for hierarchical testing and interpretability of endpoints that have not been alpha-controlled
	Missing data	Concerns around the amount of missing data (both within- and between-participants) and how missing data have been handled
	Analysis population	Impact of predefined missing data rules on the analysis sample
Interpretation of results	Inconsistency in results	Inconsistencies in outcome data
	Interpretability of results	General concerns regarding the interpretability of results, including size of between-group differences
	Issues with clinical meaningfulness	Clinical meaningfulness of results, including adequacy of evidence used to define meaningful within-patient change thresholds
Other	Concept naming	Recommendations on labeling language (e.g., how the concept is described)
	Insufficient evidence	Whether there is sufficient evidence to approve indication

Note: See supplemental table in the QR code for example FDA review language.

CONCLUSIONS

- Inclusion of diaries in clinical trials to support clinical efficacy and product labeling is generally well-accepted by the FDA.
- Diaries are being used to support labeling language across primary, secondary, and other endpoints. They are most frequently used for diseases characterized by day-to-day, event-based, or episodic fluctuations in symptoms.
- Sponsors should ensure considerations identified in this review are integrated throughout the strategic approach when evaluating the use of a diary to support a PRO label claim.

REFERENCES

1. FDA. 2023. www.fda.gov/media/168871/download
2. FDA. 2020. www.fda.gov/media/139088/download
3. FDA. 2022. www.fda.gov/media/131230/download
4. FDA. 2022. https://www.fda.gov/media/159500/download
5. Slota C, et al. Value Health. 2024;27(6):755-66.

Scan the QR code to view the full version of the poster

Contact: Sarah Sauchelli, PhD (ssauchellitoran@rti.org)

Presented at: ISPOR 2025; 13-16 May 2025; Montreal, Quebec, Canada

SUPPLEMENTAL TABLE FOR USE OF DIARIES TO CAPTURE PATIENT-REPORTED OUTCOMES IN FDA NEW DRUG APPROVALS (2019-2023)

Sarah Sauchelli,[1] Emily Chicklis,[2] Michelle K Carty,[2] Ari Gnanasakthy[3]
[1] RTI Health Solutions, Manchester, United Kingdom; [2] IQVIA, Durham, NC, United States; [3] RTI Health Solutions, Research Triangle Park, NC, United States

Supplemental Table S1. FDA Considerations When Evaluating Diary-Assessed PRO Data, With Illustrative Examples

Theme	Types of FDA feedback	Description	Example
Instrument	Fitness-for-purpose	Whether the diary or the PRO measure administered as a diary is fit-for-purpose	*"We do not have sufficient information to determine whether the PSS is a well-defined and reliable instrument that may be used to support labeling claims…provide evidence of the assessment's content validity and other measurement properties, including reliability, validity, and ability to detect change, as well as the assessment's conceptual framework and scoring in a PRO evidence dossier for FDA review and comment."* (Skyrizi/risankizumab-rzaa; plaque psoriasis; 2019)
	Comprehensiveness	Whether the diary captures all PRO concepts of importance to the target population	*"Based on discussion with Clinical, the concepts included in the modified MFSAF v2.0 are clinically relevant for the target population, with the caveat that the modified MFSAF v2.0 does not include a fatigue assessment. Fatigue is a core symptom of MF, and therefore, future studies should consider using an MF-specific assessment that includes a fatigue item(s) (e.g., the MFSAF v4). For an individual claim of fatigue improvement, sponsors should consider a separate fatigue assessment (e.g., a PROMIS Fatigue Short Form)."* (Inrebic/fedratinib; myelofibrosis; 2019)
	Device validation and implementation	Validation and implementation of the device to administer the diary measure	*"Issues identified included design and validation issues, inadequate user acceptance testing, and insufficient training of patients and study personnel on the use of the ePRO devices."* (Nurtec ODT/ rimegepant; migraine; 2020)
Study design	Overlap in endpoints	Impact of conceptual overlap in diary-assessed endpoints, in terms of patients' ability to distinguish between concepts, potential recall bias, and/or interpretability of results	*"While the assessment of multiple different abdominal symptoms has face validity, there is insufficient evidence to determine whether it is fit for regulatory purposes in terms of patients being able to differentiate among the concepts of abdominal fullness, abdominal bloating, abdominal cramping, and abdominal discomfort, and whether these concepts are considered meaningfully different from abdominal pain."* (Ibsrela/tenapanor hydrochloride; irritable bowel syndrome; 2019)
	Endpoint appropriateness	Appropriateness of endpoint, including reference to previous uses of the endpoint to support product approvals in the disease and/or for the drug	*"The primary endpoint of reduction in mean monthly migraine days over weeks 1-12 compared to placebo is the same primary endpoint used in the recent approvals for the other CGRP antagonists approved in 2018. It is also similar to the primary endpoint used in other approved preventive treatments for episodic and chronic migraine."* (Vyepti/eptinezumab-jjmr; migraine; 2020)
	Timing of administration	Whether the schedule of assessment is appropriate	*"The modified MFSAF v2.0 diary was administered at different times and the TSS was calculated using different timepoints in JAKARTA compared to COMFORT-I: […] Based on discussion with Clinical and Biostatistics, product labeling will include a statement describing the administration schedule for the modified MFSAF v2.0 diary in JAKARTA."* (Inrebic/fedratinib; myelofibrosis; 2019)
	Responder bias	Potential biases associated with diaries being patient reported	*"With the exception of overall survival, the other secondary endpoints are subject to patient and investigator bias."* (Nubeqa/darolutamide; prostate cancer; 2019)
	Single-arm study	Potential impact or mitigation of risk of bias in data collected from a single-arm study	*"The statistical reviewer concluded that the data obtained during the double-blind withdrawal period provided supportive evidence of the effect of setmelanotide on weight loss in POMC/PCSK1 populations, as it allowed each subject to serve as his or her own control."* (Imcivree/setmelanotide; obesity; 2020)
	Recall bias	Handling of potential recall bias when diary data are collected retrospectively	*"Given that the issue with not recording MBS prospectively is a concern for recall bias, examining the patients who identified MBS prospectively or within 5 minutes is a reasonable approach (since patients may be less likely to forget the symptom within a short time from taking study drug) and the results […] still appear to be valid for this endpoint in this patient population."* (Reyvow/lasmiditan; migraine; 2019)
Data quality	Unblinding	Handling of any potential unblinding and/or concerns on the impact of treatment unblinding	*"The unmasking of the syringes in 4 patients is unlikely to impact the estimated number of porphyria attacks and ALA/PBG lab results and similarly for the 2 patients who did not meet the specified inclusion criterion."* (Givlaari/givosiran; hepatic porphyria; 2019)
	Completion compliance	Concerns around completion compliance, including how it has been defined	*"There is a noted difference between the placebo, 5 mg, and 20 mg groups in regard to eCOA handheld diary noncompliance and concomitant medications deviations."* (Zavzpret/zavergepant; migraine; 2023)
	Error(s) in data entry	Inconsistencies between the data entered in an electronic diary and data entered in the case report form as well as errors identified in diary data entry	*"There were 5 subjects where the date the drug was recorded as being administered occurred after the end date of the study. The sponsor noted that the date discrepancies were due to a failure to collect the electronic diary and study drug from the patient prior to being discharged from the study. There were cases where 2 dates were listed for when a study drug was taken. The sponsor attributed this discrepancy to different dates being recorded in the electronic diary and on the case report form."* (Reyvow/lasmiditan; migraine; 2019)
Analysis	Adjustment for multiplicity	Methods used to adjust analyses for multiplicity, if any have been applied, including the sponsor's decisions on which diary-based endpoints to include for hierarchical testing and interpretability of endpoints that have not been alpha controlled	*"Secondary endpoints in the narcolepsy clinical trials were not prespecified with a plan to control for Type-I error and were considered exploratory in this analysis."* (Wakix/pitolisant; narcolepsy; 2019)
	Missing data	Concerns around the amount of missing data (both within- and between-participants) and how missing data have been handled	*"While this subsection summarizes symptom diary related efficacy endpoint data, these results should be interpreted with caution based on the following limitations… Approximately 19% subjects in the mITT2 population missed more than 25% symptom diary entries (18% in PAXLOVID group and 19% in placebo group)."* (Paxlovid/nirmatrelvir, ritonavir; COVID-19; 2023)
	Analysis population	Impact of predefined missing data rules on the analysis sample	*"The biometric review notes that when subjects with zero or missing cataplectic events were ignored, pitolisant did not demonstrate a statistically significant improvement in daily rates of cataplexy over placebo." (*Wakix/pitolisant; narcolepsy; 2019)
Interpretation of results	Inconsistency in results	Inconsistencies in outcome data	*"The fact that only 47% of patients reported no baseline pain on the BPI-SF may indicate a discrepancy in the investigator's assessment of the absence of pain compared to the patient's own description of baseline pain."* (Nubeqa/darolutamide; prostate cancer; 2019)
	Interpretability of results	General concerns regarding the interpretability of results, including size of between-group differences	*"Note that the number of phototoxic episodes may be difficult to interpret, as for an extreme example, a subject who reported a pain score of 10 every day in the study and a subject who reported a pain score of 4 on only 1 day during the study would both be recorded as having 1 phototoxic episode, even though these subjects experienced very different levels of phototoxic pain during the study."* (Scenesse/afamelanotide; erythropoietic protoporphyria; 2019)
	Issues with clinical meaningfulness	Clinical meaningfulness of results, including adequacy of evidence used to define meaningful within-patient change thresholds	*"…the clinically meaningful within-patient change threshold derived from Study 3003 was considerably higher compared with the threshold obtained from Study 008. […] Based on Study 3003 data, when you look at the aforementioned ranges, there is minimal separation between the treatment and the placebo arm."* (Germtesa/vibegron; overactive bladder; 2020)
Other	Concept naming	Recommendations on labeling language (e.g., how the concept is described)	*"For labeling, urgency should be termed 'urgency (need to urinate immediately)' as this was the term presented to patients in the PVD in studies 3003 and 3004."* (Gemtesa/vibegron; overactive bladder; 2020)
	Insufficient evidence	Whether there is sufficient evidence to approve indication	*"Confirmatory evidence of pitolisant's effect on cataplexy should be required prior to approval. Therefore, the evidence submitted with this application is not sufficient for approval of the cataplexy indication."* (Wakix/pitolisant; narcolepsy; 2020)

ALA/PBG = aminolevulinic acid/porphobilinogen; BPI-SF = Brief Pain Inventory-Short Form questionnaire; CGRP = calcitonin gene-related peptide; COVID-19 = coronavirus disease 2019; eCOA = electronic clinical outcome assessment; ePRO = electronic PRO; FDA = Food and Drug Administration; MBS = most bothersome symptom; MF = myelofibrosis; MFSAF = Myelofibrosis Symptom Assessment Form; mITT2 = modified intent-to-treat 2; ODT = orally disintegrating tablet; POMC/PCSK1 = pro-opiomelanocortin/proprotein convertase subtilisin/kexin type 1; PRO = patient-reported outcome; PROMIS = Patient-Reported Outcomes Measurement Information System; PSS = Psoriasis Symptom Score; PVD = patient voiding diary; TSS = Total Symptom Score.