PHP96

Reasons for rejection of PRO label claims: an analysis based on a review of PRO use among new molecular entities and biologic license applications 2006–2010

DeMuro C[1], Clark M[2], Mordin M[2], Evans E[1], Copley-Merriman K[2], Fehnel SE[1], Gnanasakthy A[3]

[1]RTI Health Solutions, Research Triangle Park, NC, USA; [2]RTI Health Solutions, Ann Arbor, MI, USA; [3]Novartis, East Hanover, NJ, USA

ABSTRACT

OBJECTIVES: Previous analyses of PRO label claims concentrated only on successful label claims. The goal of this research was to explore the reasons why PRO label claims were either denied or not sought.

METHODS: Using the FDA Drug Approval Report Webpage, all approved new molecular entities (NMEs) and biologic license applications (BLAs) between January 2006 and December 2010 were identified. For identified drug products, medical review sections from publicly available summary basis of approvals (SBAs) were reviewed to identify PRO endpoint status and any FDA Study Endpoints and Label Development comments.

RESULTS: Out of the 116 NMEs/BLAs identified and accompanying SBAs reviewed, 44.8% of products included PROs as part of the pivotal studies; however, only 24.1% received PRO claims. Primary reasons for denial (where data available) included a lack of demonstration of content validity (inclusive of general measures such as the EQ5D and SF-36) as well as use of PROs to assess symptoms in an open-label setting, lack of consensus on clinically meaningful change, interpretation of or missing PRO data, lack of measurement of full constellation of symptoms, issues of multiplicity and concerns of "bias" in certain PRO measures.

CONCLUSIONS: Nearly half (45%) of submissions included PROs* yet this rate is not reflected by claims granted. Understanding the nature of PRO claims granted under the current regulatory guidance is important. Additionally, a clear understanding of denied claims yields valuable insight into where sponsors may improve implementation of PROs in clinical trials and the PRO evidence submitted in order to increase the likelihood of obtaining PRO label claims.

OBJECTIVES

- Examine sponsor submissions utilizing patient-reported outcomes to explore reasons why a PRO label claim was either denied or not sought
  - Review "case studies" of claims denied
  - Categorize submission feedback to identify trends
  - Provide feedback to sponsors to increase likelihood of future claims

METHODS

- The FDA Drug Approval Reports Webpage was used to determine the number of products approved in the US from January 2006 through December 2010. Original New Drug Approvals (NDAs) and BLAs by month were selected. The reports include specification of Center for Drug Evaluation Research (CDER) NDA chemical classification. Our review included products classified by CDER as NMEs or BLAs
- Drug approval packages (DAP) and approved product labels were reviewed for each product. Information was retrieved from the medical review, summary review, cross-discipline team leader review, and other review sections from the DAP, as well as the Indication and Clinical Studies section of the approved product label according to availability. The following information was collected for each US drug product identified:
  - Brand name
  - Generic name
  - Date of approval
  - Applicant
  - Label indication
  - PRO claim language
  - PRO instruments named in label
  - Utilization of PROs
    - PROs mentioned in SBA but not appearing in the label
    - Evidence of claims sought but not granted*
    - Significance of PRO results
  - Division reviewer or SEALD reviewer feedback (where available)
  - Statistical analysis consisted of frequencies and cross-tabulations of measured characteristics. Calculations were performed using Microsoft Excel 2007

* For the purposes of this analysis, it is assumed a claim was sought unless specifically noted otherwise

RESULTS

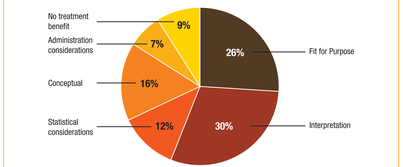
- A total of 26 products were identified as having PRO labeling claims denied across all reviewing divisions (Table 1)
  - Six products had some or partial PRO labeling granted while having other requested claims denied within the same submission
  - A wide range of PRO measures were utilized in filings, including symptom diaries, event logs, measures of functioning and disability, symptom assessments (fatigue, pain), disease-specific measures of health related quality of life, generic assessments of health related quality of life and utility measures

Table 1 Products with Claims Denied by Reviewing Division

Reviewing Division	Products Reviewed
Anesthesia, Analgesia and Rheumatology Products	CHANTIX, ILARIS
Anti-infective and Ophthalmology Products	LUCENTIS, BEPREVE
Biologic Oncology	VECTIBIX
Cardiovascular and Renal Products	LETAIRIS, SAMSCA
Dermatology and Dental Products	STELARA
Drug Oncology	DACOGEN, ZOLINZA,TORISEL, IXEMPRA KIT, TREANDA, ISTODAX, JEVTANA
Gastroenterology Products	VPRIV, ELAPRASE, RELISTOR
Medical Imaging and Hematology Products	PROMACTA
Metabolism and Endocrinology	JANUVIA, EGRIFTA, SOMATULINE
Neurology Products	AZILECT, AMPYRA
Psychiatry	INVEGA, PRISTIQ

- Primary reasons for the lack of claims were grouped for analysis purpose and included:
  - Fit for purpose – content validity, validation evidence in target population, lack of evidence of translation/cross cultural validation
  - Interpretation – issues of potential bias (open label design, etc.), recall period, clinical meaningfulness, missingness, poor compliance
  - Statistical considerations – no adjustment for multiplicity, inappropriate or missing SAP
  - Concepts – lack of link between concept and claim, failing to measure full constellation of symptoms
  - Administration considerations – lack of documentation for use of measure, copy of measure not provided to agency
  - No treatment benefit – measures did not support treatment benefit
- Over 50% of claims denied were due to issues of interpretation of PRO data (Figure 1) followed by lack of well-defined and supported evidence of "fit for purpose"
  - Specific concerns with regard to interpretation included potential bias (open label design, etc.), recall period, lack of evidence clinical meaningfulness, missingness, poor compliance with the measure (30%)
  - Issues of "fit for purpose" included content validity, validation evidence in target population, lack of evidence of translation/cross cultural validation (26%)
- Additionally, lack of support for the link from concept to claim language, statistical considerations such appropriate statistical analysis plan, plans for multiplicity adjustment, as well as lack of evidence for treatment benefit and administration issues impacted labeling decisions

Figure 1 Most Frequently Reported Reasons for Denial of Claim



- See related poster PHP97 relating to a review of PROs among NMEs and BLAs

Table 2 Case studies: PRO Claims Not Granted

Product/ Approved	Indication	Measures	PRO Feedback
AZILECT* 5/16/2006	Treatment of the signs and symptoms of idiopathic Parkinson's disease	Quality of Life Questionnaire (unnamed) Beck Depression Index Parkinson's Disease Quality of Life Scale (PD Qualif) Subject Global Improvement	Sponsor conclusions–maintenance of quality of life as compared to placebo (statistically significant improvement). Comparisons from baseline to week 26 demonstrated smaller decrement in HRQOL in treated patients as compared to placebo (significant treatment effect) Reviewer notes "I cannot draw serious conclusions about the efficacy of these endpoints because of issues of multiplicity whereby the sponsor did not make statistically appropriate adjustments for these multiple comparisons" Reviewer label recommendation: Recommend deleting presentation of any efficacy data not relevant to primary endpoints"
CHANTIX* 5/10/2006	Aid to smoking cessation treatment.	Smoking Effects Inventory	"Relief of withdrawal symptoms" deemed inappropriate claim by SEALD reviewer (due to increases in insomnia and increased appetite as noted in the AE database and slightly higher score on MMWS); use of the term "craving" questionable per SEALD (reduces urge to smoke was claim granted); SEALD reviewer noted "reinforcing effects of smoking" is not a clearly defined concept suitable for labeling
DACOGEN 5/2/2006	Treatment of patients with myelodysplastic syndromes	Quality of Life; EORTC-QLQ C30	In QOL analyses treated patients had statistically superior global health, less dyspnea and less fatigue. Reasons for lack of inclusion in labeling are not discussed in the DP
LUCENTIS 6/30/2006	For the treatment of patients with neovascular (wet) AMD and macular edema following retinal vein occlusion	Vision-Related Functioning Questionnaire-25	NEI VFQ has not been validated
ELAPRASE 7/24/2006	For patients with Hunter syndrome (can increase walking distance)	CHAQ, HS-Focus, Quality of Life assessments (by proxy)	PRO measures **did not demonstrate treatment benefit**
VECTIBIX 9/27/2006	For the treatment of metastatic colorectal carcinoma with disease progression on or following certain chemotherapy regimens	EQ-5D	Analysis of PRO data is considered **exploratory, incomplete, and potentially biased** and is not considered sufficiently robust to support a marketing claim
ZOLINZA 10/6/2006	Treatment of cutaneous manifestations in patients with CTCL who have progressive, persistent or recurrent disease on or following 2 systemic therapies	Pruritis relief: VAS, scale of 0–10	**PROs cannot be reliable measured in open-label single-arm trials**. A 3-point improvement was considered clinically significant, but the review does not state whether the proportion of patients obtaining this level of relief was clinically meaningful
JANUVIA 10/16/2006	Adjunct to diet and exercise to improve glycemic control in adults with type 2 diabetes mellitus		One study measured change in appetite using the Global Appetite/Satiety Questionnaire. **No differences in change of appetite between study drug and placebo.** This was not **used as a secondary endpoint**, and the detailed results of this measurement were not reported
TORISEL 5/30/2007	Treatment of advanced renal cell carcinoma	EQ-5D was used to measure Q-TWIPT, Caregiver Outcomes Assessment	Applicant **did not provide evidence of validation of the EQ-5D in the RCC population**. It was used in a setting for which it was not designed, and more frequently than intended. **Missing data. It was subject to bias.** The applicant also **did not provide evidence of validation of the Caregiver Outcomes Assessment**. No information from this assessment was provided in the application. **No prospective plan for the analysis of the data from caregiver reported outcomes** was revealed in the submitted statistical analysis plan
LETAIRIS* 6/15/2007	Treatment of PAH to improve exercise ability and delay clinical worsening	SF-36 Physical Functioning Scale, Borg's Dyspnea Index	**SF-36 may lack specificity for PAH. SF-36 is unevaluable b/c the preceding secondary endpoint** (WHO Functional Class) failed. Any improvement on the SF-36 score is not substantially different from an improvement of clinical worsening. "Slightly favorable effect upon dyspnea during exertion," but statistical significance has not been established
SOMATULINE DEPOT 8/30/2007	Longterm treatment of acromegalic patients who have had an inadequate response to or cannot be treated with surgery or radiotherapy	Acromegaly Symptom Measure (not named)	Studies attempted to assess the signs and symptoms of acromegaly (sweating, headache, fatigue, etc) but did not describe measures used. **Review states that the study will not allow for the conclusion that symptom improvement is a result of treatment**
IXEMPRA KIT 10/16/2007	Treatment of metastatic or locally advanced breast cancer in patients after failure of an anthracycline and a taxane	FBSI for symptomatic measurements.	**Results of PROs are difficult to interpret due to the unblinded nature of the trial and the significant drop out rate.** Statistically significant change in FBSI from baseline to week 24 favoring the capecitabine group, but it did not reach the clinically meaningful level. Review states that **no conclusions can be reached from the PRO data due to poor compliance and loss of respondents**
RELISTOR 4/24/2008	Treatment of opioid-induced constipation in patients with advanced illness who are receiving palliative care, when response to laxative therapy has not been sufficient	Evaluations of constipation distress (5-point scale), evaluations of pain (0–10 scale), evaluations of opioid withdrawal symptoms (Modified Himmelsbach, 4-point scale), patient impression of change in bowel status, patient reports of bowel consistency (6-point scale) and difficulty (5-point scale)	The label mentions that there were no clinically relevant changes in pain scores from baseline in either arm of the study. **The efficacy assessments measure constipation relief but not decrease of incidence of constipation-related complications. Measuring distress solely due to constipation in patients who are terminally ill is problematic**
SAMSCA 5/19/2009	Treatment of clinically significant hypervolemic and euvolemic hyponatremia (serum sodium < 125 mEq/L or less marked hyponatremia that is symptomatic and has resisted correction with fluid restriction), including patients with heart failure, cirrhosis, and Syndrome of Inappropriate Antidiuretic Hormone (SIADH) (1)	SF-12 Hyponatraemia Disease-specific Survey (HDS)	**Content validity of the SF-12 has not been demonstrated** for the purpose of measuring symptoms of hyponatremia in a clinical study setting to support labeling claims. The **content validity of the HDS has not been established**
ILARIS 6/17/2009	Treatment of Cryopyrin-Associated Periodic Syndromes (CAPS), in adults and children 4 years of age and older including Familial Cold Autoinflammatory Syndrome (FCAS) Muckle–Wells Syndrome (MWS) (1)	Patient assessment of symptoms. HRQoL PROs: SF-36, PCS, MCS, FACIT-F, HAD-QI in adults and CHQ-CF87 in children.	Study D2304: **HRQoL assessments did not have an adequate statistical analytic plan** to address multiplicity and should be not be included in labeling. Study A2102: **HRQoL assessments prone to bias** and should not be included. Study D2306: HRQoL data were not submitted
STELARA 9/25/2009	Treatment of adults with moderate to severe plaque psoriasis who are candidates for phototherapy or systemic therapy	DLQI, itch VAS, SF-36, HADS, Work Limitations Questionnaire, Health Economics	The DLQI **is not an adequate measure of the concept of HrQoL in the target patient population**"…does not include physical, psychological or social domains…yields overall score only…" We do not have documentation that the item generation was performed with input from the target population "…it does not appear that these items are measuring what they purport to measure [DLQI]" No justification for recall period. The SF-36 MCS and PCS **are composite measures of general health status and will not support claims of improved physical or mental functioning for labeling. The Hospital Anxiety Depression Scale is widely used, but its validity in the target population has not been established.** Health economic assessments are not intended for labeling claims. **The results of itch assessment using VAS may be suitable for labeling claims, but it is unclear what the instructions to patients were**
BEPREVE* 9/8/2009	Treatment of itching associated with allergic conjunctivitis	Ocular Itching (un-named)	Description of the clinical studies is vague and may be used by the sponsor to promote in a misleading manner. Suggest rewriting section with the following: # of patients studied in each arm of the trial, age ranges , major study endpoints, descriptions of the measurement tools used to evaluate the outcomes (the measurable signs of ocular itching), actual results (tabular format), and any appropriate accompanying statistics. Recommend that specific efficacy data be included to qualify the superiority claims made in the label. Broad claims about the superiority of the drug versus vehicle without the context of the actual data may be used to misleadingly overstate the efficacy of the drug in promotional materials. "Bepreve 1.5% was more effective than its vehicle for relieving ocular itching induced by an ocular allergen challenge, both at CAC 15 minutes post-dosing and a CAC 8 hours post dosing of Bepreve." This claim is very vague and may be used promotionally to overstate the efficacy of Bepreve. Specifically, it does not identify the specific endpoint(s) that were measured. Recommend that the claim be revised to specify the measure of relief from ocular itching to which the claim refers

Table 2 Case studies: PRO Claims Not Granted (Continued...)

Product/ Approved	Indication	Measures	PRO Feedback
ISTODAX 11/5/2009	Treatment of CTCL in patients who have had at least one systemic therapy	Pruritis VAS	**PRO data does not justify a "relief of pruritis" labeling claim. PRO data from open-label studies are rarely credible. VAS scales often produce a false sense of precision. What was measured as not relief of pruritis (as stated in proposed labeling) but rather pruritus severity at certain points in time.** There is no empirically-derived response criteria demonstrating that pruritis was actually relieved. Needs to be reviewed to make sure patients understood instructions. Useful to consider whether patients had access to their previous responses at subsequent assessments. Should be clarified whether PROs were measured at clinic visits before other clinical assessments. Some patients received medications that may have interfered with Pruritis VAS scores
AMPYRA* 1/22/2010	To improve walking distance in patients with MS	MSWS-12 Subject Impression of Change (SGI)	**MSWS-12 walking scale has not been fully validated to support a disability claim, full validation would be required (in collaboration SEALD) Subject Impression of Change (SGI) was no better for fampridine than for placebo (p=0.122).** SGI was evaluated by asking patients to rate themselves based on the following question: "how do you feel about the effects of the study medication over the past7 days?", on a scale 0-7, where 0 was "terrible" and 7 "delighted". **The lack of significant difference on that endpoint questions the clinical relevance of the effect noted on the responder rate and the MSW12**
JEVTANA 6/17/2010	In patients with hormone refractory metastatic prostate cancer previously treated with a docetaxel-containing regimen	Pain intensity--McGill--Melzack Present Pain Intensity scale	**Open-label trials are not adequately designed to support efficacy** conclusions based on subjective PRO measures such as pain intensity. **Validity of the data is questionable**. Other deficiencies were identified that suggests that pain intensity was not adequately measured. **Content validity of the measure has not been adequately established. No justification that pain progression is clinically meaningful. Trial inclusion criteria did not include baseline pain intensity criteria**. Use of the morphinic equivalent table in determining an analgesic score is problematic b/c it's unclear similar analgesic scores produce the same degree of analgesia for all patients. **Sponsor has not included a copy of the analgesic log used.** No statistically significant difference in pain progression between the treatment arms. **The largest number of protocol violations involved missing pain assessments or analgesic scores—216 missing pain assessments in 106 patients**
EGRIFTA* 11/10/2010	Reduction of excess abdominal fat in HIV-infected patients with lipodystrophy	Body Image Impact Module (BIIM)	Study Endpoints and Label Development (SEALD) consult. Body Image Impact Module (BIIM), reviewer expresses **reservations with respect to the content validity** of the current version of the BIIM, which does not meet the new standards articulated in the December 2009 guidance and states that it should not be recommended by FDA for future drug development. Of note, the PROs evaluated in the Egrifta clinical trials have been incorporated with input from the Agency much in advance of the 2009 guidance. The reviewer notes that Belly Size Estimation is **not a valid measure of "belly size" as it asks the subject to compare his/her current belly size to his/her ideal "healthy look" and thus, in the absence of more specific criteria, it is doubtful that the term "healthy look" will be interpreted the same way across subjects and even for the same subject over time.** The reviewer also indicates that **Belly Appearance Distress may be a valid measure but the data provided in the PRO dossier does not meet the new standards for instrument development** recommended within the 2009 PRO guidance. The consult defers to the statistical and clinical team the judgment as to whether the data submitted and critically evaluated methodologically in the consult can be considered clinically meaningful and adequate for labeling. In response to the comments and recommendations made by the SEALD consult, the clinical and statistical team decided to include in the label only the results of the Belly Appearance Distress. It was felt that from a clinical perspective Belly Appearance Distress is an endpoint of higher significance as it does not measure the self-reported perception about changes in the size of the abdomen but rather the emotional impact and distress for the patient, an important proxy for QOL in HIV-patients with lipodystrophy
INVEGA	For the acute and maintenance treatment of schizophrenia, schizoaffective disorder	Symptoms and Quality of Life in Schizophrenia (SQLS) Sleep VAS	Proposed product labeling will contain only a primary endpoint and 1 key secondary endpoint. Any data for other secondary endpoints will not be reviewed No adjustments for multiple comparisons were employed for group comparisons on the VAS sleep scores. There were statistically significant result for both SQLS and Sleep VAS t
PRISTIQ	Treatment of major depressive disorder	VAS Pain Intensity	VAS-PI was not the key secondary efficacy variable. FDA disagreed with sponsor's proposed list of key secondary outcome measures and indicated that most could not be used in product labeling. FDA noted that they would accept 1 measure of the CGI scale as a key secondary endpoint
TREANDA	For treatment of patients with Chronic lymphocytic leukemia (CLL)	Quality of Life as measured by EORTC-C30 and EORTC QLQ-CLL25	Results of QoL measurements were not reported in the DAP
PROMACTA	Treatment of thrombocytopenia	Incidence and severity of symptoms associated with chronic ITP, SF-36	Results of PRO measures were not reported in the DAP
VPRIV	Enzyme replacement therapy (ERT) for pediatric and adult patients with type 1 Gaucher disease	SF-36 (for adults) CHQ, PF50 (for patients 5-17 years of age)	Limited availability of QoL data because measures were not used in all trials. No QoL conclusions due to small number of patients

*products had some claims granted

- Table 2 provides a detailed analysis of regulator feedback on PRO submissions by product, as available, in the drug approval packages
  - Differing levels of detail/information were provided in the drug approval packages and review formats were inconsistent
- Discrepancies in the acceptability of measures across product reviews are noted
  - For example, use of the SF-36 physical component score was allowable in labeling for Savella (Mordin et al 2011) but not for Stelara
- Products not receiving label claims were frequently used in peer-review publications in support of product communication strategies

CONCLUSIONS

- PRO label claims are denied for various reasons – some of which are addressed by the FDA PRO guidance
- Review of claims denied yields valuable insight into where sponsors may improve implementation of PROs in clinical trials and the level of PRO evidence submitted in order to increase the likelihood of obtaining PRO label claims

LIMITATIONS

- Due to the confidential nature of label discussions with the FDA, data may have been withheld from the SBA
- Sponsor intent may not have been appropriately interpreted
  - For purposes of analysis, we assumed a label claim was sought (unless otherwise specifically stated)
  - The sponsor may not have intended to seek label claims (PRO data may have been included as supportive efficacy information and as part of a communication strategy) influencing the level of evidence submitted to the agency
  - Some studies designed and executed prior to release of the draft guidance
- Drug approval packages yielded differing levels of information and detail in regard to PROs
  - Not all received specific SEALD review
  - Specific PRO evidence dossiers were not available for review

REFERENCES

1. US Department of Health and Human Services. Draft Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. February 2006.
2. US Department of Health and Human Services. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. December 2009. Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. Accessed January 14, 2011.
3. Willke RJ, Burke LB, Erickson P. Measuring treatment impact: a review of patient-reported outcomes and other efficacy endpoints in approved product labels. Control Clin Trials. 2004 Dec;25(6):535-52.
4. Mordin M, Clark M, DeMuro C, Evans E, Copley-Merriman K, Fehnel S, Gnanasakthy A. PRO label claims: an analysis based on a review of PROs among NMEs and BLAs 2006-2010. ISPOR 16th Annual International Meeting, May 21-25, 2011.

Poster presented at ISPOR 16th Annual International Meeting, May 21–25, 2011, Hilton Baltimore Baltimore, MD, USA