

Going Pro with PROs: Implementation and Analysis of Patient Reported Outcomes

Shilpakala Vasudevan, Efficacy Lifescience Analytics

ABSTRACT

In recent times, Patient Reported Outcomes (PRO) have been gaining a lot of popularity as significant endpoints, in analyzing a subject's condition in a clinical trial. PRO data are collected directly from the subjects without any interpretation from the investigator. Collected Information is generally related to how well they are able to perform functions in daily life, their health related quality of life (HRQOL), and satisfaction with the health care provided. For collecting and reporting this information, suitable instruments need to be identified, that are comprehensible, reliable and easy to work on. Many questionnaires are available that record responses on pain/discomfort, mobility, anxiety/depression, and the physical, functional, emotional and social well-being of a patient. Examples of such questionnaires include EQ-5D, FACT-P, EORTC QLQ-C30, and SF-36. This paper will cover the aspects of collection, data handling, statistical analysis, and reporting of PROs, with examples of datasets and statistical procedures used.

INTRODUCTION

Patient Reported Outcomes are gaining a lot of popularity in recent times. In patient reported outcomes, data is collected directly from the patient, with which they can record their thoughts, feelings, opinions and grievances. It can thus provide insights on a patient's perspectives on treatment and outcomes. Patient reported outcomes (PROs) typically measure the ease with which daily functions are performed, symptoms that exist, and the patient's health related quality of life.

To collect data from patients, we need instruments or tools that will enable us to record their responses. There are many such questionnaires already widely in use that we can choose from. In addition, new questionnaires could also be developed, that needs to have a conceptual framework and that could also accommodate the endpoints of the study. The ADQS dataset is the widely used dataset for patient reported outcome analysis. PROs are generally collected over a period of time as longitudinal data, and analyses like change from baseline and between treatment comparisons are then performed for further interpretation.

CLINICAL OUTCOME ASSESSMENTS

Clinical Outcome Assessments (COA) are used in clinical trials to measure a patient's functioning on different levels, like measuring their symptoms, how they feel on an overall state, or to determine the effects of a condition on the patient. COAs can be used to verify if there are treatment benefits on the condition of patients and are hence used to define endpoints in studies.

There are 4 types of Clinical Outcome Assessments:

1) Clinician-Reported Outcome (ClinRO) --

When the concept can be observed by a trained healthcare professional and requires clinical judgement by a clinician. For example, studying the performance of a patient is done by clinicians.

2) Performance Outcome (PerfO) --

When the concept can only be observed, by the patient performing a set of defined tasks in a clinical setting, as instructed by health care professionals. PerfOs could be physical in nature like checking the grip strength or walking speed, or cognitive in nature like testing the memory.

3) Observer-Reported Outcome (ObsRO) --

When the concepts can only be observed outside a healthcare setting, and where the patient is not able to make the recording of the observation by themselves. For example, where patients do not have the cognitive ability to record, the observer can observe and make the reports.

4) Patient-Reported Outcome (PRO) --

When the concept can only be observed by the patients and where the patients are able to take note of their observations. Example of PROs are noting the symptoms, and the frequency or severity of symptoms.

PATIENT REPORTED OUTCOMES

As per the FDA, a Patient Reported Outcome (PRO) is any report of the status of a patient's health condition that comes directly from the patient, without any interpretation of the patient's response by a clinician or anyone else. Medical technology enables us to measure physiological and biochemical data of patients, but there is some information which cannot be physically measured. Such data can only be collected from the direct experience of the patients. In many cases, only a patient will be able to perceive some important factors which cannot be detected by any other person or technical instrument.

Implementing Patient Reported Outcomes in a study is recommended when a concept is best measured from a patient's perspective. It could additionally be used to support existing data, and to find the outcomes of specific treatments and to prove their effectiveness. With patient feedback, PROs can help understand what patients are undergoing, their feelings and how to bring about the desired results for them. Overall, PROs if made an integral portion of clinical research in any therapeutic area, could help in improving a patient's life as well improving lives overall for the entire population in a study.

PRO vs. Proxy Reported Outcomes

Sometimes, when patient reported outcomes are not available, proxy reports are accepted for questions that can be easily observed. The measurements based on these proxy reports by some other person, is considered as proxy reported outcomes. These are different from patient reported outcomes and observer reported outcomes.

PRO vs. Patient-Centered Outcomes (PCO)

Patient-Centered Outcome (PCO) are based on reports that cover issues and concerns that are specific to a patient. Patient-reported outcomes (PRO) are different in that, information is provided by only the patient, is related to only the treatment or condition, and the information in the report may or may not be of concern to the patient.

PRO vs. Patient-Reported Experience Measures (PREM)

Patient-Reported Experience Measures (PREM) report on a patient's overall experience while they were receiving treatment care. PREMs measure experience as in for example, communication and timelines of assistance received while under care. Patient-reported outcomes (PROs) on the other hand, measure the outcomes of care and treatment.

COLLECTING PATIENT REPORTED OUTCOMES

Different instruments and tools exist to capture data from patients in a clinical trial, that are called as patient reported outcome measures (PROMs). PROs are many times synonymously called as PROMs. A PRO instrument typically consists of questionnaires, with the information and documentation along with it to support in its use.

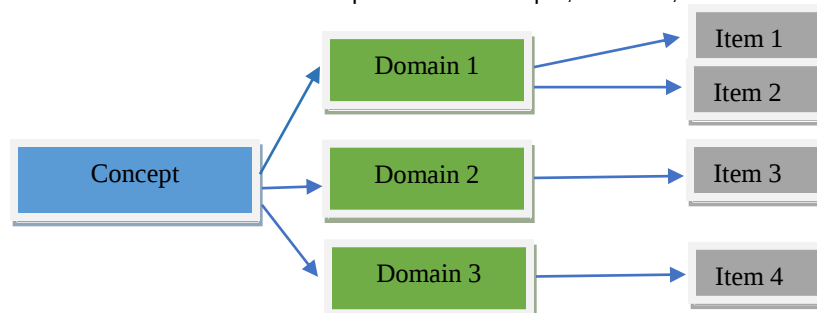
To determine the instrument that is to be used, first the objectives and endpoint model for the study should be defined. Patient Reported Outcomes are widely used as endpoints these days. Depending on the study, they could be considered as primary, secondary or exploratory endpoints. In studies, for example, where the treatment of the disease is the primary concern and the alleviation of symptoms are secondary concerns, PROs could be secondary endpoints. If the disease symptoms are of primary concern, they could be considered as primary endpoints.

Endpoints and endpoint model

Based on the endpoint model, we could determine the kind of PRO instrument we would like to use in our study. First, we determine whether an adequate PRO instrument already exists that could be used to measure the concepts and derive the endpoint. If the instrument is not fully compatible for our requirement, we could work on modifying the instrument. If modification will also not be suitable, a new PRO instrument will need to be developed.

Conceptual framework

The instrument that is used by the study, should have a conceptual framework designed for it. A conceptual framework defines the relationship between concepts, domains, and items as can be seen in the following diagram:



PhUSE US Connect 2018

Concept – The event that is measured in the study, by the instrument. For example, measuring symptoms could be the concept.

Domain – A sub concept that is measured within the concept. For example, fatigue and nausea could be the domains.

Item – A question that is evaluated by a patient. For example, measuring the severity and frequency of fatigue or headache could be items.

The number of domains and items depend on the condition or disease being treated. If the concept of interest is only measuring intensity of pain, then a single item instrument may be enough. However, if the concept is general in nature, as like in measuring the overall functioning of a patient, then it would involve multiple domains and multiple items.

When new PRO instruments are developed, they need to be evaluated on the following considerations:

- 1) Validity – The extent to which the instrument measures the concept of interest
- 2) Reliability – Ability of an instrument to yield consistent estimates and stable scores
- 3) Ability to detect change – The ability for the instrument to detect changes in outcomes over time

All items in the questionnaire must be relevant to the patients in the study and should be addressing the actual condition. They should not be hypothetical in nature and should not include questions on functions the patients are not likely to perform. Questions with less responses may suggest very little activity and produce a bias towards the null, showing no effect of treatment.

Items should be worded correctly and should be clear and readable.

The responses in the item could be in one of the following formats. The responses should be well ordered with appropriate intervals, and should be suitable for the population in the study:

What is the severity of your pain?

Visual Analog Scale (VAS): 0 _____ 100

What is the severity of your pain?

Anchored or Categorized VAS: 0 20 40 60 80 100

What is the severity of your pain?

Likert Scale: ☐ No pain ☐ Moderate pain ☐ Extreme pain

How do you rate your pain on scale of 1 (low) to 5 (high)?

Rating Scale: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Rate the severity of your pain

☐  ☐  ☐ 

Pictorial Scale: No pain Moderate Extreme

Is your pain severe?

Checklist: ☐ Yes ☐ No

Also, using not applicable as an option for the questions is not suggested, as it will be difficult for analysis.

The items and instructions for the questionnaire should be clear to read and understandable.

Recall period (time the patient can validly recall information) for the item responses should be appropriately decided.

Modes of administration and data collection

Administration of questionnaires could be self-done by the patient or could be done by interview or combination of both.

Electronic modes of data collection are the most preferred as they help in regular timed reports from the patients' home, and real-time monitoring can be done by the investigators.

Telephone based approach is suitable when patients have limited functional abilities or not enough literacy.

Paper-based reporting is least preferred as compliance could be less and monitoring would not be too easy.

PhUSE US Connect 2018

SCORING AND ANALYSIS

Once the responses in the questionnaire are recorded and collected from the patient, measurement is done by scoring all the items. First, numerical scores are assigned for each answer category in every item. These scores are assigned on an appropriate scale of measurement for the item.

The scores are then combined for all items together to form a single score for that domain. Similarly, scores from all domains would be then computed to get the single score for the concept in consideration. Computing the common score is determined by a scoring algorithm for the questionnaire. This would also take into consideration the appropriate weights that need to be allocated to each item.

Questionnaires

There are many questionnaires available that are widely used, tested and validated. Questionnaires can be categorized into two types:

i) Generic Questionnaires address generic concepts that can be applicable to patients in general and is not directed towards any particular disease. Below are some of the generic questionnaires available:

- EuroQol-5D (EQ-5D)
- 36-Item Short Form Health Survey (SF-36)
- European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30)
- e-Health Impact Questionnaire (eHIQ)
- Health Utilities Index (HUI)

ii) Disease specific questionnaires are applicable to only the disease or condition in consideration in the study.

The following are a few examples of disease-specific questionnaires:

- Musculoskeletal Health Questionnaire (MSK-HQ)
- The Nutrition and Dietetic Patient Outcomes Questionnaires (NDPOQ)
- Diabetes Health Profile (DHP)
- The Myocardial Infarction Dimensional Assessment Scale (MIDAS)
- Functional Assessment of Cancer Therapy (FACT)
- Brief Pain Inventory (BPI)
- Rheumatoid Arthritis Quality of Life (RAQoL)

Below are a few examples of using some questionnaires, the scoring process and analyses:

EXAMPLE 1: SINGLE ITEM QUESTIONNAIRE

Eastern Cooperative Oncology Group (ECOG)

Performance status data are collected in many oncological studies to determine how well they can perform daily functions. Usually, ECOG is clinician reported and is considered gold standard when collected by the clinical person. However, when done by a clinician, it may introduce bias as it is from a different person's perception and it is subject to misinterpretation. IN ECOG, there are 6 scales with 0 indicating complete functioning down to 5 indicating death of a patient.

ECOG is a single item questionnaire, as there is only question that is recorded for response. Below is an annotated CRF for ECOG:

ECOG PERFORMANCE STATUS		QSCAT=ECOG
<u>Grade</u>	<u>ECOG</u>	
☐ 0	Fully active, able to carry on all pre-disease performance without restriction	
☐ 1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work	
☐ 2	Ambulatory and capable of all selfcare but unable to carry out any work activities. Up and about more than 50% of waking hours	
☐ 3	Capable of only limited selfcare, confined to bed or chair more than 50% of waking hours	
☐ 4	Completely disabled. Cannot carry on any selfcare. Totally confined to bed or chair	
☐ 5	Dead	QSORRES when QSTESTCD = ECOG101
		QSSTRESQ/QSSTRESN

PhUSE US Connect 2018

ADQS is the commonly used ADAM dataset for questionnaires. PARAM and PARAMCD contain the item from the questionnaire, and AVAL is the score recorded.

usubjid	trt01p	parcat	Param	paramcd	avisit	adt	aval	avalc
01-1001	Treatment A	ECOG	ECOG Performance Status	ECOG101	Baseline	02JUN2016	2	Ambulatory and capable of all selfcare
01-1001	Treatment A	ECOG	ECOG Performance Status	ECOG101	End of Treatment	16MAY2017	4	Completely disabled
01-1002	Treatment B	ECOG	ECOG Performance Status	ECOG101	Baseline	17SEP2012	3	Capable of only limited selfcare
01-1002	Treatment B	ECOG	ECOG Performance Status	ECOG101	End of Treatment	09MAY2014	5	Dead

The performance status can then be compared, on a before and after treatment basis. A Table with frequencies for ECOG status can be seen below:

Summary of ECOG Performance Status, pre and post chemotherapy

	Treatment A (N=100)		Treatment B (N=98)		Total (N=198)	
	Pre-chem n (%)	Post-chem n (%)	Pre-chem n (%)	Post-chem n (%)	Pre-chem n (%)	Post-chem n (%)
0	0	0	1 (1.0)	0	1 (0.5)	0
1	15 (15.0)	10 (10.0)	14 (14.3)	11 (11.2)	29 (14.6)	21 (10.6)
2	22 (22.0)	19 (19.0)	23 (23.5)	20 (20.4)	45 (22.7)	39 (19.7)
3	29 (29.0)	31 (31.0)	33 (33.7)	25 (25.5)	62 (31.3)	56 (28.3)
4	24 (24.0)	27 (27.0)	21 (21.4)	34 (34.7)	45 (22.7)	61 (30.8)
5	0	2 (2.0)	0	1 (1.0)	0	3 (1.5)
Not reported	10 (10.0)	11 (11.0)	8 (8.2)	9 (9.2)	18 (9.1)	20 (10.1)

EXAMPLE 2: SINGLE DOMAIN MULTIPLE ITEM QUESTIONNAIRE

Fatigue Severity Scale (FSS)

Fatigue Severity Scale is a single domain instrument, that measures the domain of fatigue alone. There are 9 questions in FSS measuring fatigue, and response is recorded from 1 (Strongly disagree) to 7 (Strongly agree).

Here is a sample of 3 questions from Fatigue Severity Scale CRF:

Fatigue Severity Scale (FSS) QSCAT=FSS

Please circle the number that indicates your degree of agreement with each statement below, where 1=strongly disagree and 7=strongly agree.

1. My motivation is lower when I am fatigued.
QSTESTCD=FSS0101

2. Exercise brings on my fatigue.
QSTESTCD=FSS0102

3. I am easily fatigued.
QSTESTCD=FSS0103

Strongly Disagree

Strongly Agree

QSORRES

1

2

3

4

5

6

7

1234567

1234567

QSTRESSC/

QSTRESSN

PhUSE US Connect 2018

As seen below in the ADQS dataset, param changes with every item (question), but the parcat (domain) remains the same.

usubjid	trt01p	parcat	param	paramcd	avisit	base	aval	chg
02-1011	Treatment A	FSS	Motivation	FSS0101	Baseline	6	6	
02-1011	Treatment A	FSS	Motivation	FSS0101	Cycle 1 Day 1	6	3	3
02-1011	Treatment A	FSS	Exercise	FSS0102	Baseline	4	4	
02-1011	Treatment A	FSS	Exercise	FSS0102	Cycle 1 Day 1	4	3	1
02-1011	Treatment A	FSS	Fatigue	FSS0103	Baseline	5	5	
02-1011	Treatment A	FSS	Fatigue	FSS0103	Cycle 1 Day 1	5	1	4

Change from baseline is another common kind of analyses performed on patient reported outcomes. Change from baseline (chg) is computed in ADQS.

Change from Baseline for Fatigue Severity Scale

Baseline Fatigue Severity Score	Change from Baseline Fatigue Severity Score						
	1	2	3	4	5	6	7
	n	n	n	n	n	n	n
1	31	2	0	0	0	0	0
2	19	15	2	0	0	0	0
3	16	11	7	1	1	0	0
4	12	9	4	0	2	1	1
5	7	4	0	6	3	0	0
6	3	2	2	0	0	7	1
7	5	5	3	2	9	5	10

EXAMPLE 3: MULTIPLE DOMAIN SINGLE ITEM QUESTIONNAIRE

EuroQol-5 Dimension (EQ-5D)

EuroQol-5D is a generic health status instrument, that measures the health state in 5 domains: mobility, self-care, usual activities, pain/discomfort, anxiety/depression. The EQ-5D-5L has 5 responses for every question, and EQ-5D-3L has 3 responses for every question. Besides the 5 domains, there is a visual analogue scale EQ VAS, that the patient records their health status with, on a scale of 0 (worst imaginable health state) to 100 (best imaginable health state).

Under each heading, please check the ONE box that best describes your health TODAY

Mobility QSORRES when QSCAT =EQ5D01

I have no problems in walking about ☐

I have some problems in walking about ☐

I am confined to bed ☐

Self-Care QSORRES when QSCAT =EQ5D02

I have no problems with self-care ☐

I have some problems washing or dressing myself ☐

I am unable to wash or dress myself ☐

Usual Activities (*eg, work, study, housework, family or*

leisure activities) QSORRES when QSCAT =EQ5D03

I have no problems with performing my usual activities ☐

I have some problems with performing my usual activities ☐

I am unable to perform my usual activities ☐

PhUSE US Connect 2018

Pain/Discomfort **QSORRES when QSCAT =EQ5D04**

I have no pain or discomfort ☐

I have moderate pain or discomfort ☐

I have extreme pain or discomfort ☐

Anxiety/Depression **QSORRES when QSCAT =EQ5D05**

I am not anxious or depressed ☐

I am moderately anxious or depressed ☐

I am extremely anxious or depressed ☐

Since this involves many domains, we need a common overall score that can be used to determine the overall health status. For this purpose, a scoring algorithm and published weights are available along with the questionnaire, that can be used to calculate the overall score. The overall scores calculated range from 0 to 1, with 0 representing worst health, and 1 representing perfect health.

The coefficients (weights) have been published and are available and differ for each country. Following is an example of coefficients as measured by the quality adjusted life years in UK. The coefficient for every question varies as per the response.

Domain	Raw score	Response	coefficient
Mobility	1	No problems	
	2	Some problems	-0.069
	3	Confined to bed	-0.314
Self-Care	1	No problems	
	2	Some problems	-0.104
	3	Unable	-0.214
Usual Activities	1	No problems	
	2	Some problems	-0.036
	3	Unable	-0.094
Pain/Discomfort	1	No pain	
	2	Moderate pain	-0.123
	3	Extreme pain	-0.386
Anxiety/Depression	1	Not anxious	
	2	Moderately anxious	-0.071
	3	Extremely anxious	-0.236

Following ADQS dataset has responses with values of 1,2,3 for every question, and appropriate coefficient (weight) is included for every response:

usubjid	trt01p	parcat	param	paramcd	aval	avalc	coefficient
03-1111	Treatment A	EQ5D01	Mobility	EQ5D0101	2	Some problems	-0.069
03-1111	Treatment A	EQ5D02	Self-Care	EQ5D0201	2	Some problems	-0.104
03-1111	Treatment A	EQ5D03	Usual Activities	EQ5D0301	1	No problems	
03-1111	Treatment A	EQ5D04	Pain/Discomfort	EQ5D0401	2	Moderate pain	-0.123
03-1111	Treatment A	EQ5D05	Anxiety/Depression	EQ5D0501	3	Extremely anxious	-0.236

PhUSE US Connect 2018

As per the EQ-5D algorithm, the total health index is calculated by adding up the coefficients, and subtracting from 1.
 EQ-5D Health State = 1 – (Coefficient1 + Coefficient2 + Coefficient3 + Coefficient4 + Coefficient5)

So, in this case:

$$\text{EQ-5D Health State} = 1 - (0.069 + 0.104 + 0 + 0.123 + 0.236) = 0.468$$

Since 0 indicates worst health and 1 indicates best health, a value of 0.468 falls somewhere close to moderate health.

usubjid	trt01p	parcat	param	paramcd	aval
03-1111	Treatment A	EQ5D	Health State	EQ5DHLST	0.468

Comparison of health states between treatment groups can be performed two-sample t tests as below:

```
PROC TTEST data=adqs;
  by avisit;
  class trt01p;
  var aval;
run;
```

EQ-5D – Observed Means and Change from Baseline

	Treatment A (N=100)				Treatment B (N= 98)				Treatment A – Treatment B		
	n	Mean(Std)	Median	95% CI	n	Mean(Std)	Median	95% CI	Mean	95% CI	P-Value
Mobility											
Observed Value											
Cycle 1 Day1 (Baseline)	100	0.67	0.63	(0.60,0.74)	98	0.53	0.52	(0.50,0.56)	-0.12	(-0.10,-0.14)	<0.001
Cycle 2 Day1	97	0.85	0.83	(0.80,0.90)	96	0.65	0.64	(0.62,0.68)	-0.17	(-0.12,-0.22)	<0.001

Note 1: P-values are based on 2-sample t-test

Note 2: Higher scores are associated with better health status and lower scores for worse health.

EXAMPLE 4: MULTI DOMAIN MULTI ITEM QUESTIONNAIRES

Functional Assessment of Cancer Therapy: General (FACT-G)

FACT-G is used to assess health-related quality of life of cancer patients, and also of patients that undergo a variety of other chronic illnesses. FACT-G has 4 domains (Physical, Social, Emotional and Functional well-being), with 6-7 questions (items) in each domain.

EMOTIONAL WELL-BEING

		Not at all	A little bit	Some-what	Quite a bit	Very much
GE1	I feel sad	0	1	2	3	4
GE2	I am satisfied with how I am coping with my illness.....	0	1	2	3	4
GE3	I am losing hope in the fight against my illness.....	0	1	2	3	4
GE4	I feel nervous.....	0	1	2	3	4
GE5	I worry about dying.....	0	1	2	3	4
GE6	I worry that my condition will get worse	0	1	2	3	4

FUNCTIONAL WELL-BEING

		Not at all	A little bit	Some-what	Quite a bit	Very much
GF1	I am able to work (include work at home)	0	1	2	3	4
GF2	My work (include work at home) is fulfilling.....	0	1	2	3	4
GF3	I am able to enjoy life.....	0	1	2	3	4
GF4	I have accepted my illness.....	0	1	2	3	4
GF5	I am sleeping well	0	1	2	3	4
GF6	I am enjoying the things I usually do for fun	0	1	2	3	4
GF7	I am content with the quality of my life right now.....	0	1	2	3	4

PhUSE US Connect 2018

The scores from the questionnaire above are first recorded as raw scores.

A raw score of 0 may indicate a lower health or higher health, depending on the question. So, to get to a common scale, of lower value of 0 as least health score and the highest value as the best health score, we need to first transform the raw scores to z-scores, by performing reversals. Reversal is performed by adding or subtracting values given in the instruction manual, for each question.

For example, for emotional well-being, to get the reversed score, for all questions we subtract the values from value 4, except for the 2nd question:

FACT0301: 4-3=1 FACT0302: 1 FACT0303: 4-4=0

FACT0304: 4-2=2 FACT0305: Missing FACT0306: 4-1=3

usubjid	trt01p	parcat	param	paramcd	aval	z-score
04-1001	Treatment A	FACT03	I feel sad	FACT0301	3	1
04-1001	Treatment A	FACT03	I am satisfied with coping	FACT0302	1	1
04-1001	Treatment A	FACT03	I am losing hope	FACT0303	4	0
04-1001	Treatment A	FACT03	I feel nervous	FACT0304	2	2
04-1001	Treatment A	FACT03	I worry about dying	FACT0305		
04-1001	Treatment A	FACT03	I worry my condition will get worse	FACT0305	1	3

After reversal, we sum all z-scores in the domain, and multiply by the number of items in the domain (6) and divide by number of items answered (5). This produces the T-score for every domain.

$$\text{FACT03} = (1+1+0+2+3)*6 / 5 = 8.4$$

Below is the T-score for the domain Emotional well-being for this subject:

usubjid	trt01p	parcat	param	paramcd	T-score
04-1001	Treatment A	FACT03	Emotional well-being	FACTSCR	8.4

All domain scores are added to get the total FACT-G score.

FACT-G score = Physical T-score + Social T-score + Emotional T-score + Functional T-score

The lower the score, the lower the QoL, and the higher the score, the better the quality of life.

Below is an analysis based on mixed effect model:

FACT-G— Between Treatment Comparison

	Treatment A (N=50)		Treatment B (N=47)		Treatment A – Treatment B		
	Estimated Mean	95% CI	Estimated Mean	95% CI	Estimated Mean	95% CI	P-Value
Emotional Functioning							
Post Baseline Observations:							
Cycle 2 Day 1	78.3	(65.0,90.0)	82.1	(74.6,88.3)			
Cycle 3 Day 1	73.6	(63.3,82.9)	77.5	(68.9,85.8)			
Overall Comparisons of Post Baseline Observations:							
Treatment	75.4	(62.9,87.1)	80.3	(71.8,83.9)	4.9	(3.3,6.5)	0.04
Time							0.03
Treat x time interaction							0.036

Note: Analysis based on repeated measures mixed effects model with treatment, time, treatment-by-time interaction as fixed effects. P-values are from Type 3 Tests of Fixed Effects.

EXAMPLE 5: TIME TO EVENT FOR QUALITY OF LIFE

Time to Deterioration (TTD) of quality of life

Time to deterioration is the time it takes for the quality of life scores to deteriorate with a minimal clinically important score change, from the baseline score. For example, deterioration could be considered if there is a 10 points or higher increase in cough symptom for two consecutive cycles. In this case time to deterioration (TTD) is calculated from baseline to the time when 10 points or higher increase is detected continuously for two cycles.

PhUSE US Connect 2018

Below is the ADTTE (ADaM Time to Event) dataset for Time to Deterioration (TTD) :

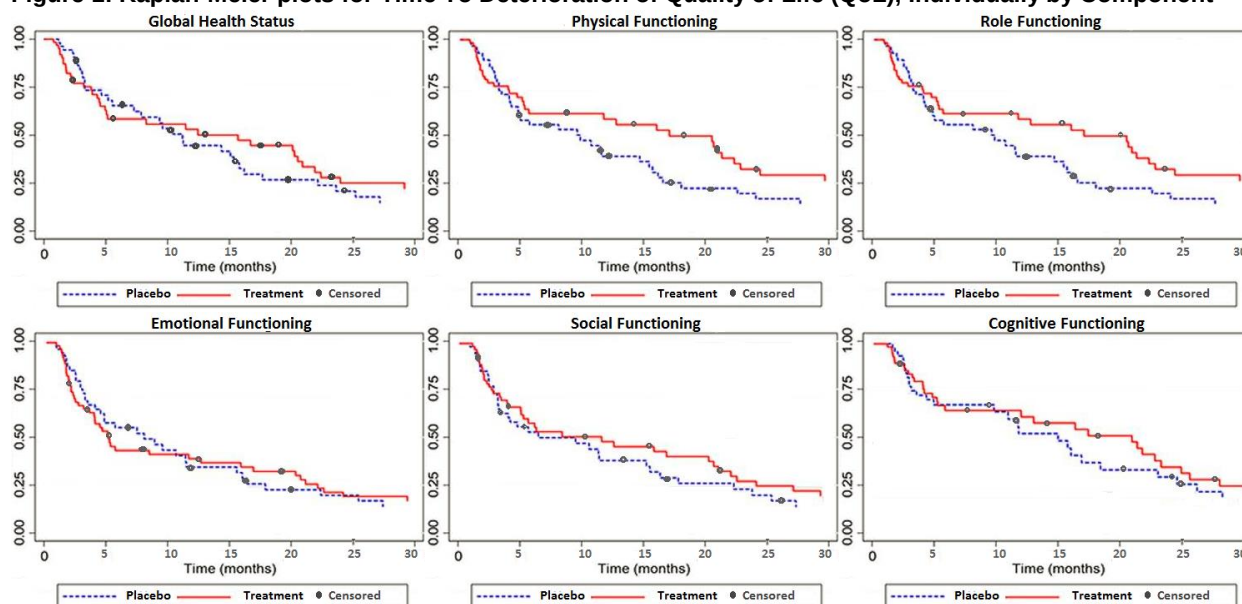
usubjid	param	paramcd	randdt	adt	aval	cnsr	evntdesc	cnsdtc
05-1011	Time to deterioration of QoL (Days)	TTD	10FEB2016	07NOV2017	637	2	Withdrawal by Subject	Date of last contact
05-1012	Time to deterioration of QoL (Days)	TTD	25JUN2013	28APR2017	1404	1	End of study	Date of last contact
05-1013	Time to deterioration of QoL (Days)	TTD	17SEP2012	12JUN2016	1365	0	Deterioration	
05-1014	Time to deterioration of QoL (Days)	TTD	09MAY2014	01MAY2017	1089	3	Subject Discontinuation	Date of last contact
05-1015	Time to deterioration of QoL (Days)	TTD	29JAN2015	28FEB2017	762	0	Deterioration	
05-1016	Time to deterioration of QoL (Days)	TTD	30OCT2016	14JAN2017	77	4	Lost to Follow-up	Date of last contact

Statistical analyses using Kaplan-Meier estimates are most common for time to event analyses:

```
PROC LIFETEST data=adtte method=km outsurv=dsplot;
  time aval*cnsr(1);
  by trtpn trtp;
run;
```

The Kaplan-Meier plots can be generated either for every symptom (like cough, pain, nausea) or it can be generated for every component of Health Related Quality of Life (Global Health Status, Physical Functioning, Role Functioning, Emotional Functioning, Social Functioning and Cognitive Functioning), as shown below:

Figure 1: Kaplan-Meier plots for Time To Deterioration of Quality of Life (QoL), Individually by Component



EXAMPLE 6: SAFETY PATIENT REPORTED OUTCOMES

Patient Reported Outcomes-Common Terminology Criteria for Adverse Events (PRO-CTCAE)

Adverse Event data is traditionally collected by clinicians. PRO instruments too can be used to get safety information from the patients to include their perspectives and improve safety reporting. Patient Reported Outcomes of Adverse-Events (PRO-AEs) are based on a patient-centric approach and are used when concerns like signs or symptoms are best obtained from the patient.

PhUSE US Connect 2018

There are many advantages to using PRO-AE data, along with clinician reported AE data.

- The patients will be able to observe any symptoms in a better manner than the clinicians.
- Also, patient-reporting will enable frequent visits which will help in better longitudinal analyses.
- Reporting done with National Cancer Institute Common Terminology Criteria for Adverse Events (CTC[AE]) for symptom capture depend on the clinician to correctly take the data to a medical record and then to data entry. When done without proper training, this could lead to possibilities of error, inaccuracies in reporting and lack of appropriate interpretation of the symptoms.
- In oncological trials, CTCAE baseline is often not reported by clinicians. Baseline symptoms can be easily detected by the patients. With the help of a baseline patient-reported AE, we will be easily able to identify if the symptoms were pre-existing or are treatment emergent.

The National Cancer Institute (NCI) Patient Reported Outcomes-Common Terminology Criteria for Adverse Events (PRO-CTCAE) is a measurement tool that is used recently to identify the characteristics of 78 different symptoms in oncology. There are 124 questions available that measure the presence, amount, frequency, severity and interference with usual or daily activities for the symptoms. A 5-point Likert scale is used to record responses and the recall period is the previous 7 days.

Below is a sample of a symptom in PRO-CTCAE questionnaire:

22. PRO-CTCAE™ Symptom Term: Swelling				
a. In the last 7 days, how OFTEN did you have ARM OR LEG SWELLING?				
☐ Never	☐ Rarely	☐ Occasionally	☐ Frequently	☐ Almost constantly
b. In the last 7 days, what was the SEVERITY of your ARM OR LEG SWELLING at its WORST?				
☐ None	☐ Mild	☐ Moderate	☐ Severe	☐ Very severe
c. In the last 7 days, how much did ARM OR LEG SWELLING INTERFERE with your usual or daily activities?				
☐ Not at all	☐ A little bit	☐ Somewhat	☐ Quite a bit	☐ Very much

The ADAE dataset will record the Clinician and Patient reported Adverse Events with their toxicity grades:

usubjid	trt01p	aereport	aedecod	aestdt	aetoxgr	aetoxgrn
06-1111	Treatment A	Clinician	Nausea	02MAR2015	Grade 1	1
06-1111	Treatment A	Patient	Nausea	24FEB2015	Grade 3	3
06-1112	Treatment B	Clinician	Nausea	25MAY2016	Grade 0	0
06-1112	Treatment B	Patient	Nausea	02MAY2016	Grade 2	2

When reporting this data, there are two options. Clinician reported AEs that are only non-symptomatic can be presented with patient reported AEs that are only symptomatic. Or, both clinician-reported and patient-reported AEs can be combined and presented in a single table as shown below:

Summary of Any Grade of Adverse Events, Clinician-reported and Patient-reported				
		Treatment A	Treatment B	p-value
		(N=100)	(N=100)	
		n (%)	n (%)	
Nausea	Clinician-reported	20 (20)	16 (16)	0.22
	Patient-reported	26 (26)	22 (22)	0.12
Abdominal pain	Clinician-reported	35 (35)	21 (21)	0.54
	Patient-reported	42 (42)	25 (25)	0.33
Cough	Clinician-reported	10 (10)	0	0.15
	Patient-reported	13 (13)	9 (9)	<0.001
Swelling	Clinician-reported	22 (22)	15 (15)	0.32
	Patient-reported	29 (29)	17 (17)	0.24
Palpitations	Clinician-reported	32 (32)	28 (28)	0.17
	Patient-reported	37 (37)	33 (33)	<0.001

Note: p-value from Fisher's exact test, comparing grades between treatments

PhUSE US Connect 2018

MISSING DATA

One of the main challenges with patient reported outcomes is the occurrence of missing data. Missing data greatly reduces precision due to reduction in data, and could potentially lead to bias, and make it difficult to compare the effects between two treatment groups due the reduced set of population.

- Data that is missing is indicative of either the patient not understanding how to complete the questionnaire or that the items were not clear or meaningful to them. It could also mean that there were lost survey pages in the questionnaire or that the electronic surveys were wrongly programmed. One way to counter this, is to get early information of the correctness of the instrument by reviewing the percentage of missing scores, in addition to other descriptive statistics.
- Additionally, missing data could arise because patients sometimes fail to report for visits or withdraw from a clinical trial before its completion. It must be ensured that PRO measurements be obtained immediately before or after withdrawal of the patient. Close monitoring, setting up reminders to complete the questions and using electronically reported questionnaires need to be implemented to ensure compliance.

In some questionnaires, imputation could be applied to the missing data, depending on the reason for the missing data and the missing pattern.

Imputation for missing data may be done, by either single or multiple imputation.

- Single imputation is the simplest method where a missing value is replaced with the last observed value for that variable, or with the worst score. However, this can underestimate standard errors, confidence intervals and p-values.
- Multiple imputation replaces missing observations with random values from a statistical model and creates several imputed datasets. Analysis is then done on each of the imputed datasets to obtain estimates, and the final estimate is the average of the estimates of all the datasets. This is the preferred method of imputation for missing values.

CONCLUSION

Patient reported outcomes are very valuable in assessing overall benefits of a treatment in recent times. Measuring survival is an important feature in most cancer studies. While longevity is an important goal, we also need to make sure prolonged survival is not acquired at the cost of poor life quality for the patients. Hence, patient reported outcomes are widely included in oncological studies. Although commonly used in oncology and psychological studies, it is recently widely used for other therapeutic areas as well.

Challenges can be avoided by having a proper study design and reducing questionnaire complexity. One challenge is the difficulty patients may have in recalling events of longer periods. These can be helped with having frequent reports from the patients and forming questions that limit recall to a shorter period. Also missing data can be avoided to a certain extent by giving clearer instructions and with better follow-up. Additionally, complexities in scoring can be reduced by transforming PRO measures to a unidirectional standardized scale and then generating a single overall score.

REFERENCES

Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims, December 2009. <https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf>

CDISC "Questionnaires, Ratings and Scales (QRS)" <https://www.cdisc.org/foundational/grs>

"FACIT Questionnaires" www.facit.org/facitorg/questionnaires

"EuroQoL Questionnaires" <https://euroqol.org/eq-5d-instruments/>

Ethan Basch, Lauren J. Rogak, and Amylou C. Dueck, 2016 "Methods for implementing patient-reported outcome (PRO) measures of symptomatic adverse events in cancer clinical trials" Clinical Therapeutics

CONTACT INFORMATION

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

Shilpakala Vasudevan
Ephicity Lifescience Analytics
Email: shilpakala.vasudevan@ephicity.in

Brand and product names are trademarks of their respective companies.