Using real-world data to generate reference values for patient-reported outcome measures

Kim Cocks[1], Bryan Bennett[1], Elaine Brohan[1], Katherine Byrne[2], Alex Rider[2]
[1]Adelphi Values, Bollington, Manchester, UK; [2]Adelphi Real World, Bollington, Manchester, UK

BACKGROUND

PROs: The interpretation problem

> Patient-reported outcomes (PROs) are increasingly used in clinical trials and real-world studies to provide both direct and indirect evidence of treatment benefit by assessing how patients feel or function (e.g. symptoms, HRQoL) [1].
> Interpretation of PRO scores can be challenging due to different measurement scales and unfamiliarity with scores from new instruments. Instrument-specific guidelines are useful for interpretation.

Reference values: part of the solution

> Reference values relating to specific disease subgroups or the general population provide context to aid interpretation of such PRO scores e.g. scores obtained by lung cancer patients can be compared with those of a matched control group of persons with a different cancer or with the general population.
> The European Organisation for Research and Treatment of Cancer (EORTC) group further highlight five main areas within which reference values are beneficial: 1) comparisons of a group of patients with similar characteristics e.g. to explain differences in clinical outcomes, such as progression or death; 2) to increase familiarity with the distribution of scores for a scale; 3) sample size calculation; 4) comparison of an individual patient's score with patients with similar characteristics and 5) quality control in translation procedures.[2]
> Reference values are normally displayed within disease subgroups by age category and gender. To compare with a specific study population you would generate the age and gender matched reference values using a simple weighted equation e.g.

Study population	Mean physical functioning (study population)	Reference value males with same disease	Reference value females with same disease	Weighted reference value	Interpretation
40:60 males: females	63.0	65.4	70.2	(0.4 x 65.4) + (0.6 x 70.2) = 68.3	Study population has a lower than average physical functioning compared to the reference population (63.0 vs 68.3)

Reference values: the evidence gap

> Large sample sizes are required across a heterogeneous group of people to achieve reliable reference values, e.g. the Queensland cancer risk study which obtained general population reference values for the Functional Assessment of Cancer Therapy General (FACT-G) involved completing n=9419 telephone interviews[3] and data were included from n=2,236 patients in Brucker et als' study[4] to obtain US disease-specific reference values for the FACT-G. The time consuming nature of this work means that there is a lag time between publication of a new PRO and availability of reference values e.g. the FACT-G instrument was published in 1993[5] with first reference values published in 2004 for cancer survivors[6] and then 2005 for the US population.[4] The EORTC-QLQ-C30 was first published in 1993[7] with first reference values available in 1997[8]
> Furthermore, reference values may be unavailable for many regions where PRO instruments are used. In our example we required reference values for the FACT-G for a European population and these were not available.[9]
> Where reference values are available, in some cases reports were conducted more than 10 years ago, and updated reference values reflecting current PRO values are necessary.[10]
> It should be noted that the FACT-G and EORTC QLQ-C30 are well validated and established PROs with a number of sources of reference values, however this is often not the case and the need for reference values is likely to be greater for less well known and utilized instruments.

Real world evidence: Filling the gap

> This poster highlights the potential of real-world data for use in generating reference values using the Adelphi Real-World Disease Specific Programmes™ (DSPs) as an example. Real world databases contain large numbers of patients, normally across a range of ages, genders and disease stage/characteristics so can be a good basis for the estimation of reference values and provide a timely alternative to a large prospective study.
> The example shows estimation of reference values from an EU sample of cancer patients for the FACT-G instrument[5] using real world data. Reference values based on a US population are already published[4], we aimed to see if the DSP data could generate reference values based on an EU population and highlight the strengths and limitations.

METHODS

Data Source

> The Adelphi Real World DSPs™ provide data from independent, cross-sectional surveys conducted across Europe (France, Germany, Italy, Spain, UK – the 5EU) between January 2015 and March 2017.
> 4,899 patients across the 5EU countries completed the FACT-G measure as part of the DSP™. Data were gathered across breast, gastric, melanoma, non-small cell lung (NSCLC) and prostate cancers. Patients' treating physicians (n=1,534) provided corresponding clinical and treatment related variables.
> The full DSP™ methodology has been utilized in over 50 disease areas, is outlined in Figure 1 and is published.[11]

PRO: FACT-G

> The FACT-G contains four domains including physical well-being (PWB), social/family well-being (SWB), emotional well-being (EWB), and functional well-being (FWB). PWB, SWB, and FWB scores range from 0-28 and EWB 0-24.
> All questions are rated on a 5-point Likert-type scale ranging from 0 (not at all) to 4 (very much). The FACT-G score is the sum of the four general subscales and has a maximum score of 108. Higher scores represent better functioning and QoL.

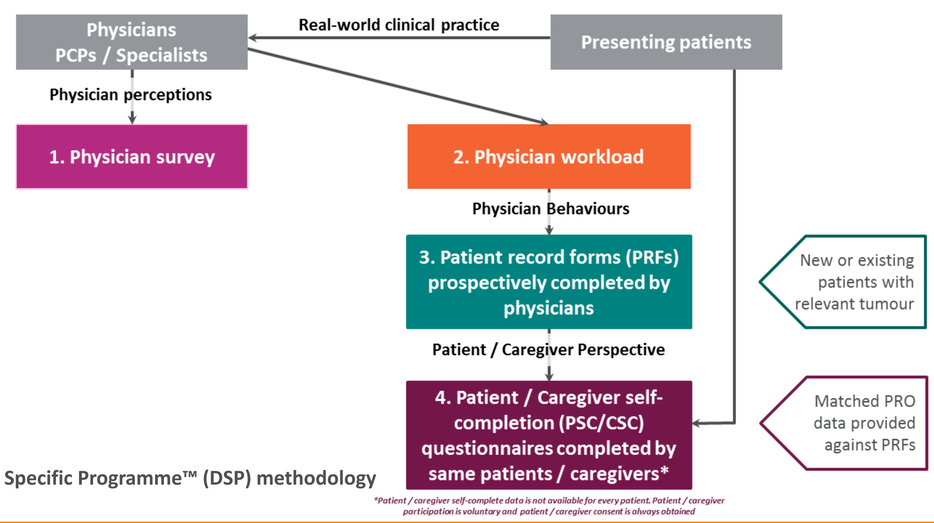

Figure 1: The Disease Specific Programme™ (DSP) methodology

RESULTS

Sample characteristics

> The 5EU real world (RW) sample had different population characteristics to the existing US reference value sample. In particular the sample covered a different range of cancer types (EU data excluded haematological cancers but had data from gastric, melanoma and NSCLC). In the 5EU sample, with the exception of prostate cancer where 15%(n=178) of patients were classified as stage I-II, all other patients across all tumour types were classified as having advanced diseases (stage III-IV).
> The demographic and clinical characteristics for the two cohorts are presented in Table 1, Figures 2 and 3.

Table 1: Demographics for 5EU and US data from Brucker et al.

Demographics	Group	Frequency (%)	
		5EU RW (n=4,899)	US Brucker (n=2,236)
Age Group	18-34	47 (1%)	135 (6%)
	35-44	207 (4.2%)	286 (12.8%)
	45-54	587 (12%)	494 (22.1%)
	55-64	1436 (29%)	607 (27%)
	>=65	2616 (53%)	711 (32%)
	Missing	6 (0%)	3 (0%)
Gender	Female	2113 (43%)	1271 (57%)
	Male	2786 (57%)	965 (43%)
Ethnicity	Afro – Caribbean	74 (1.5%)	-
	Asian – other	18 (0.4%)	-
	Asian – Indian subcontinent	13 (0.3%)	-
	Chinese	16 (1.2%)	-
	Hispanic/Latino	95 (1.9%)	69 (3.1%)
	Middle Eastern	60 (1.2%)	-
	Missing / Not available	1715 (35%)	-
	Mixed race	38 (0.8%)	-
	Other	9 (0.2%)	21 (0.9%)
	White/Caucasian	2861 (58.4%)	1534 (68.6%)
	Non-hispanic Black	-	610 (27.3%)
	Missing	-	2 (0.1%)

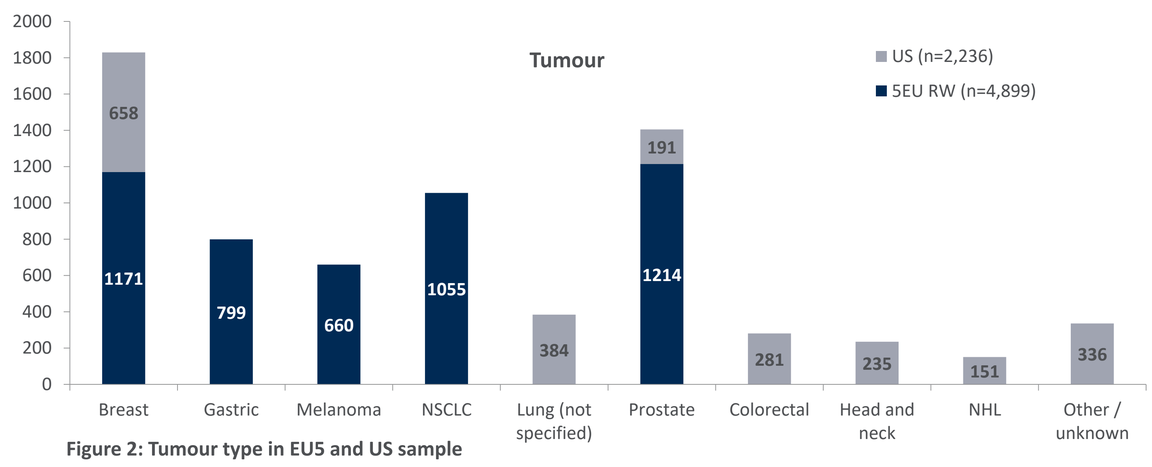

Figure 2: Tumour type in EU5 and US sample

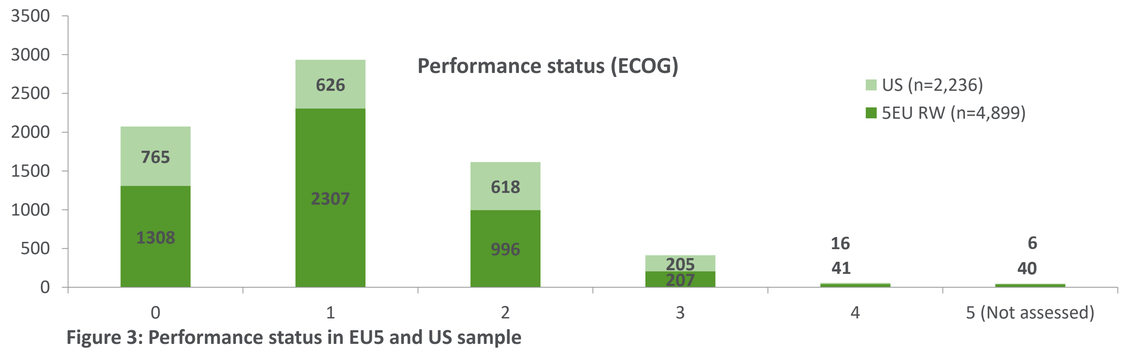

Figure 3: Performance status in EU5 and US sample

5EU RW and US reference values

> The EU generated reference values are presented below alongside the previously published US values for comparison. Figures 4, 5 and 6 display the mean scores for the whole sample and for males and females respectively.

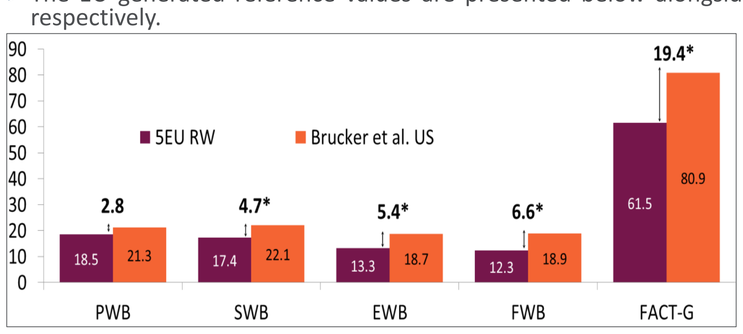

Figure 4: Difference between mean scores for 5EU RW data and Brucker et al. US data

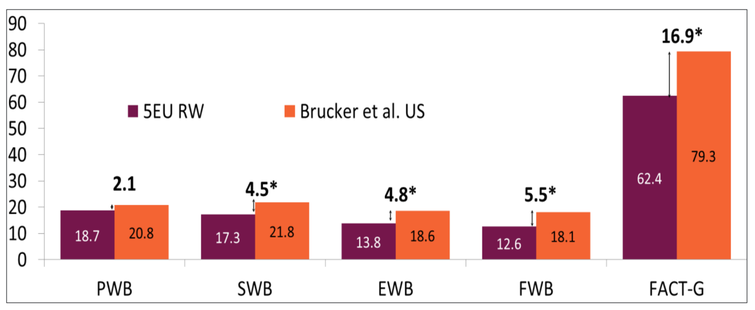

Figure 5: Difference between mean scores for males 5 EU RW data and Brucker et al. US data

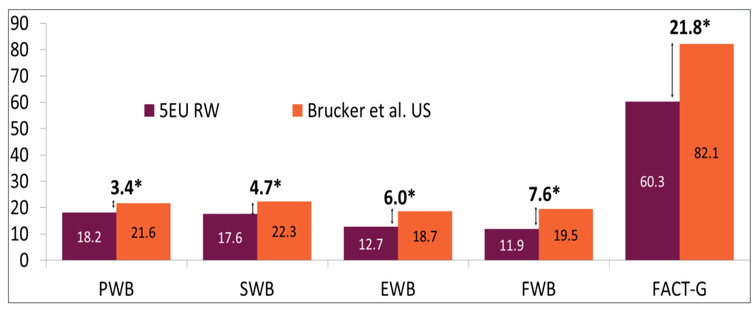

Figure 6: Difference between mean scores for females 5 EU RW data and Brucker et al. US data

CONCLUSIONS AND LIMITATIONS

> RWE where PRO data has been captured can be used to generate reference values since the sample sizes are large and will generally be sufficient for this purpose.
> They could be a fast way of generating reference values where none have been published for a specific country or disease sub-type.

Limitations

> Reference values from different sources may have a different case mix and therefore will not be directly comparable. Differences in reference values between US and EU here are probably due to the difference in patient sample rather than true differences between the regions. Ideally reference values would cover the full range of disease sub-types but this may not be possible from existing RW databases.

REFERENCES are available separately

Presented at PSI ANNUAL CONFERENCE, Beurs Van Berlage, Amsterdam, 4 June 2018