PLOS ONE

RESEARCH ARTICLE

A quantitative method for measuring the relationship between an objective endpoint and patient reported outcome measures

Chul Ahn[1☯], **Xin Fang**[1☯*], **Phyllis Silverman**[1‡], **Zhiwei Zhang**[2‡]

1 Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, Maryland, United States of America, **2** Department of Statistics, University of California, Riverside, CA, United States of America

☯ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* xin.fang@fda.hhs.gov

Citation: Ahn C, Fang X, Silverman P, Zhang Z (2018) A quantitative method for measuring the relationship between an objective endpoint and patient reported outcome measures. PLoS ONE 13 (10): e0205845. https://doi.org/10.1371/journal.pone.0205845

Editor: Stefano Marchetti, University of Pisa, ITALY

Received: June 10, 2018

Accepted: October 2, 2018

Published: October 25, 2018

Data Availability Statement: Data are available from the Harvard Dataverse at: https://doi.org/10.7910/DVN/LWQBSF.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Patient reported outcome measures (PROMs) become increasingly important for assessing the effectiveness of a drug or medical device. In order for a PROM to be claimed in labeling, the PROM has to be valid, reliable and able to detect a change if the targeted disease status changes. One approach to assess the quality of a patient reported outcome measure (PROM) is to investigate the association between the PROM and an objective clinical endpoint measuring the status of a disease/condition. However, methods assessing the association between continuous and discrete variables are limited, especially for correlated measurements. In this paper, we propose a method to assess such association with any type of samples with or without correlation. The method involves estimating the probability revealing the status of a subject's disease/condition (called truth thereafter) through the subject's reported outcomes. The probability is a conditional probability revealing truth given the relative location of the subject's objective outcome compared to the subject-specific latent threshold in the objective endpoint. A consistent estimator for the probability is derived. The operating characteristics of the consistent estimator are illustrated using simulation. Our method is applied to hypothetical clinical trial data generated for an ophthalmic device as an illustration.

1. Introduction

Patient reported outcome measures (PROMs) have become increasingly important in measuring the effectiveness of a drug or medical device. Between years 1997 and 2002, about 30% of the new drug labels were found to have included patient reported outcomes (PROs) [1]. Later between 2006 and 2010, about 24% of new molecular entities and biologic license applications were granted patient reported outcome (PRO) claims [2]. The authors of this paper also noticed that the PROM claims in approved medical devices had been steadily increasing since 2012. In the meantime, many efforts have been made to advance the use of PROMs in drug or

medical device development and regulatory decision making. Recent major challenges were reported from the Food and Drug Administration's perspective [3]. The National Institutes of Health (NIH) also funded the establishment of a PROM Information System (PROMIS) [4, 5, 6]. Some recent literature focuses on the interpretation of PRO analysis results [7, 8].

In order for a PROM to be claimed in labeling of a drug or medical device, the PROM has to be valid, reliable and able to detect a change if the status of the targeted disease or condition changes [9]. The most frequently and broadly used statistics in a PROM validation such as Pearson and intra-class correlation coefficient (ICC) [10] assess the association among PROM items or between a PROM and other established measurement(s). These correlation coefficients have been used to examine various validities (e.g. construct, convergent/divergent, criterion) of PROMs [10–30]. The correlation coefficients were also used to investigate the PROM's ability to detect a change [31]. Some authors also used these correlation coefficients to explore the relationship of a PROM with other measurements [32–35].

However, these correlation coefficients (1) may not be appropriate in correlated samples such as repeated measures, (2) may not be reliable for endpoints with different scales (e.g. categorical scale vs. continuous scale), and (3) do not have an intuitive clinical meaning because these coefficients or their changes don't directly carry a clinical meaning. It is difficult to draw a line for an acceptable association based on these popular coefficients most likely due to the lack of clinical meaning of these correlation coefficients.

The challenge here is to develop a meaningful reliable methodology to measure the relationship between an objective continuous endpoint (X) and the dichotomized endpoint (G) of an ordinal PROM, and if the association index is strong enough, to use only the PROM to make inference about the effectiveness of the therapy or to use the PROM to support the primary inference in a clinical trial setup. This paper provides such a new meaningful quantitative statistic measuring the conditional association (denoted as Q here after) between paired endpoints (X, G), and a method to translate the ordinary PROM scales to the continuous objective measurement. The use of conditional association is due to the fact that the outcome of G is conditional on the outcome of X, because the PROM is always administrated after the treatment takes effect. The dichotomized endpoint (G) may represent mixed Bernoulli random variables with the same parameter but opposite meaning, which is explained in the method section of this paper.

Section 2 describes the definition of the conditional association parameter Q, the data structure used in this paper and how to estimate Q. The derivation of the estimator of Q is also presented in this section. Section 3 shows simulation results of the estimator ($\hat{Q}$) of Q and an application of this new methodology to hypothetical clinical trial data. The discussion and conclusion are presented in Section 4.

2. Methods

This section shows how the parameter Q works in assessing the quality of a PROM using repeated measures from a single subject. It starts with minimum notations and theoretical construct of Q, followed by the characteristic and estimation procedure of Q, and the derivation of the consistent estimator of Q. The derivation of the estimator is specifically arranged after introducing the estimation process so that the derivation is more accessible to readers. The section ends with how to obtain the inference for the PROM in multiple subjects.

In general, a single italic lower-case letter represents a nonrandom variable and a single italic upper-case letter represents a random variable unless stated otherwise (such as parameter Q). The non-italic PROM_z (not a random variable) represents the scale z of the unidimensional PROM. Q_{iz} is the probability of the PROM_z revealing the disease status of Subject i according

to his/her latent minimum objective threshold a_{iz} given the subject's objective outcome $x_i \geq a_{iz}$ or $x_i < a_{iz}$. Note: Q_{iz} is not defined as a random variable and is a parameter to be estimated. The italic *PROM* is the random variable for the subjective PRO measurement, and the italic *PRO* is the realization of the *PROM*. The italic PRO_z represents the patient reported outcome equal to the scale z of the PROM. Other notations are defined in Appendix A.

2.1 Theoretical construct of parameter Q_{iz}

As illustrated in Fig 1 below, the theoretical construct of Q_{iz} is that there is a latent minimum threshold a_{iz} of a disease status in terms of the objective disease measurements of Subject i which triggers PRO_z ($z = 1, \ldots, 7$ in Fig 1) upon the PROM question according to the association parameter Q_{iz} given the subject's objective outcome $x_i \geq a_{iz}$. Although the PROM question and scales don't change with subject, sub-index i is used to indicate that the $PROM_i$ is the PROM random variable for Subject i, hereafter for clearance and without a loss of generality. Subject i will give his/her $PROM_i \geq z$ with probability Q_{iz} when his/her $x_i \geq a_{iz}$, and will give his/her $PROM_i < z$ with probability Q_{iz} when $x_i < a_{iz}$. Note here, the PRO_i is always dependent on where the X_i is realized relative to the minimum latent threshold a_{iz}.

Fig 1 illustrates the relationship between the continuous objective endpoint X_i (such as increase in hemoglobin count (HC)) and a unidimensional 7-scale $PROM_i$ (such as fatigue

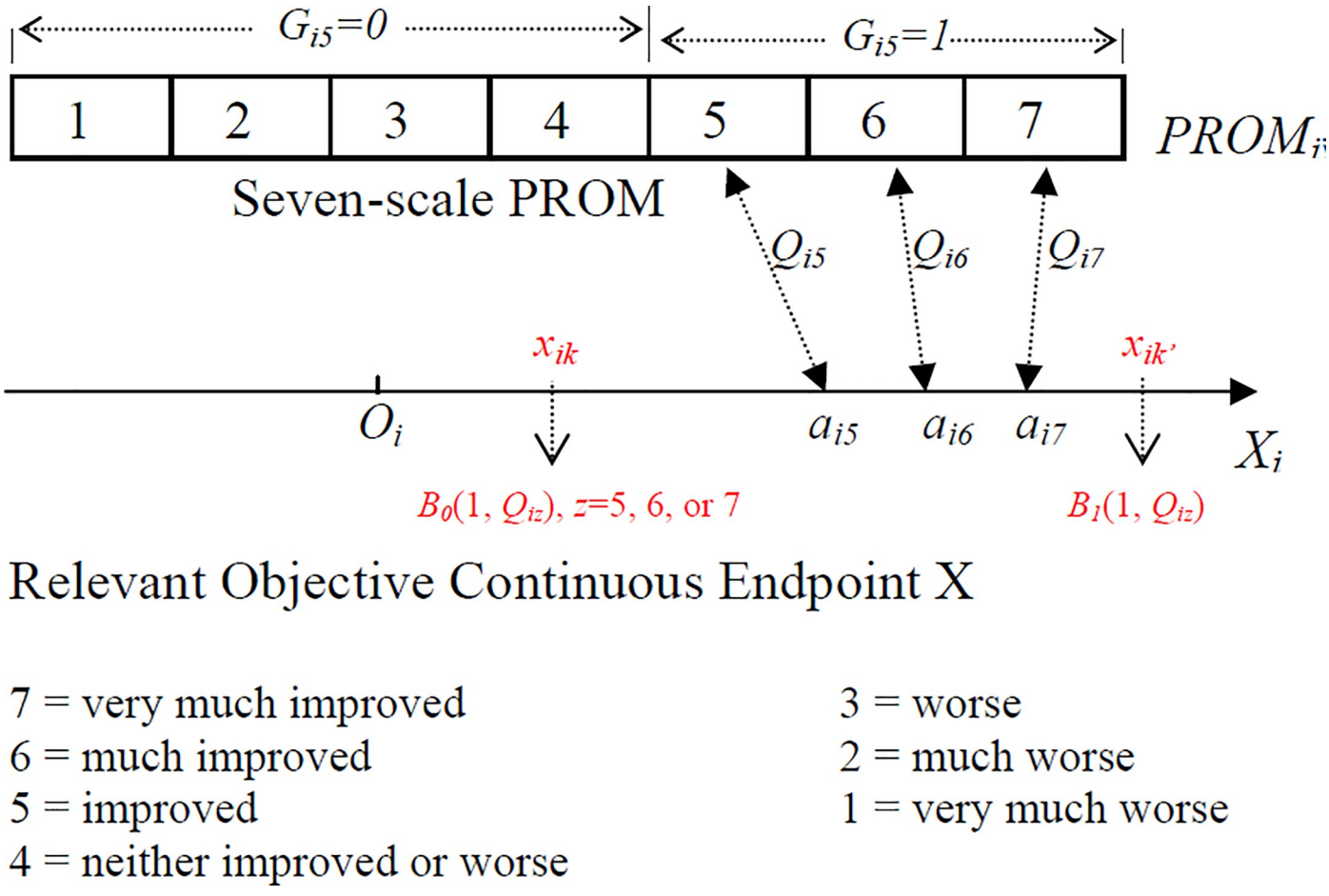

Fig 1. Conditional associations between a *PROM* and a continuous objective efficacy endpoint *X* for subject *i*.

https://doi.org/10.1371/journal.pone.0205845.g001

improvement). The upper divided rectangular block illustrates a 7-scale unidimensional PROM, and the lower line X illustrates the continuous objective measurement with letter *O* indicating the baseline location of a subject. Each scale of the PROM (such as 5 = improved) for Subject *i* has its own minimum latent objective threshold (such as a_{i5}) pointed by a connecting arrow between the two measurements. The *PROM* will be realized to the PRO_z with probability Q_{iz} by Subject *i* upon the PROM question if $x_i \geq a_{iz}$, which determines the conditional association of the $PROM_i$ with the continuous objective endpoint X_i for Subject *i* at $PROM_z$.

Note here, the event of "$PROM_z$" revealing the disease status of Subject *i* includes two true events: (1) $PROM_i \geq z$ if $x_i \geq a_{iz}$ (as true positive), and (2) $PROM_i < z$ if $x_i < a_{iz}$ (as true negative). We realize that if there is no conditional association between $PROM_i$ and X_i both $Pr(PROM_i \geq z \mid X_i \geq a_{iz})$ and $Pr(PROM_i < z \mid X_i < a_{iz})$ are equal to the pure chance rate: 50%. Therefore, we are searching the minimum threshold a_{iz} in this paper such that Subject *i* will give his/her $PROM_i \geq z$ with probability Q_{iz} when $x_i \geq a_{iz}$; and likewise Subject *i* will give his/her $PROM_i < z$ also with probability Q_{iz} when $x_i < a_{iz}$. If the probability between the two possible "truths" are not equal, their estimations require many more assumptions (see derivation section for details) and are not considered in this paper. It is also necessary to point out that the two probabilities are not complementary to each other.

2.2 Characteristics of parameter Q_{iz}

Parameter Q_{iz} varies with $PROM_z$ and subject based on its definition. Therefore, there is no linear relationship between the *PROM* and the objective endpoint *X* for any subject. For example, $Pr(PROM_i < 5 \mid X_i < a_{i5})$ may be different from $Pr(PROM_i < 6 \mid X_i < a_{i6})$; and $Pr(PROM_i < 5 \mid X_i < a_{i5})$ for Subject *i* may be different from $Pr(PROM_h < 5 \mid X_h < a_{h5})$ for Subject *h*.

It is obvious that the clinical meaning of Q_{iz} is inherited from its definition; i.e. **the rate of revealing the truth, conditional on disease status** (the actual disease status of Subject *i* relative to his/her minimum latent objective threshold for $PROM_z$). A 50% rate revealing truth is equivalent to the subject flipping a fair coin to determine his/her PRO_z upon the PROM question; thus, this rate of 50% revealing truth indicates that the $PROM_z$ is not able to reveal the subject's disease status. In general, the higher the rate revealing truth is, the better the quality of the $PROM_z$ is. This is because the higher rate indicates a higher probability of the $PROM_z$ to reveal a subject's disease status upon the PROM question.

The use of Q_{iz} to reveal the actual status of a subject's disease has not been discussed in literature. Rasch promoted a probability model for a true positive response [36]. However, because a negative agreement was not considered, the Rasch positive probability did not measure the probability of revealing truth from a PROM. Our approach is related to latent variable models for similar problems [37, 38] in the sense that a_{iz} can be regarded as a latent variable. On the other hand, we do not assume a particular distribution for a_{iz}, which makes our approach different from most latent variable models. It is also noteworthy to know that Q_{iz} is also measuring an indirect agreement between a continuous endpoint and a dichotomized version of an ordinal endpoint. Most traditional methodologies for measuring agreement as described in [39] are developed for two measures of the same type: both categorical or both continuous endpoints. In the case of different types of endpoints, ranks within each endpoint will replace the original values to make the two endpoints the same type (such as Spearman CC). In addition, the estimation of Q_{iz} (1) can be applied to correlated data, (2) takes into consideration the uncertainty of the "gold standard" and involves a series of 2-by-2 tables in order to select one for the estimate (see the toy example below). Therefore, Q_{iz} can be also viewed as a new agreement statistic between a continuous endpoint and a binary endpoint with or without correlation among samples.

2.3 Data and corresponding random variables

The data considered in this paper consist of pairs of observations (x_{ik}, g_{ik}) for Subject i at clinical visit k, where $k = 1, \ldots, t$. This x_{ik} is a continuous outcome representing disease status and could be the value at visit k or the change from baseline to visit k, such as the change in hemoglobin count from baseline. The outcome g_{ik} is the dichotomized version of the collected *PRO*s at visit k, such as $g_{ik} = 1$ if the $PROM_i \geq 5$ and $g_{ik} = 0$ otherwise. The change from baseline in the $PROM_i$ is not considered here, because (1) each latent threshold of a $PROM_z$ is corresponding to the $PROM_z$ itself instead of its change, and (2) a change in *PRO*s from baseline does not carry the same clinical meaning, which depends on the baseline *PRO*s. For example, in a 7-point scale $PROM_i$ shown in Fig 1, a change in one PROM unit from "much worse" to "worse" may not be meaningful to a subject, while a change in one PROM unit from "neither" to "improved" carries clinical meaning to the subject.

The corresponding random variables are denoted as (X_{ik}, G^1_{ik} or G^0_{ik}). The G^1_{ik} is the Bernoulli random variable ($B_1(1, Q_{iz})$) with probability Q_{iz} to be 1 when $x_{ik} \geq a_{iz}$, and G^0_{ik} is the Bernoulli random variable ($B_0(1, Q_{iz})$) with parameter Q_{iz} to be 0 when $x_{ik} < a_{iz}$. In other words, upon the PROM question, Subject i will give his/her $g_{ik} = 1$ (positive) with probability Q_{iz} when his/her $x_{ik} \geq a_{iz}$, and will give his/her $g_{ik} = 0$ (negative) with probability Q_{iz} when his/her $x_{ik} < a_{iz}$ as illustrated in Fig 2 below.

2.4 Estimation of Q_{iz}

This subsection shows how to estimate Q_{iz} using a toy example. The derivation of the estimator of Q_{iz} can be found in next subsection. In order to estimate Q_{iz}, it is necessary to first search a_{iz}. Because the a_{iz} is the minimum latent threshold in the objective measurement for the $PROM_z$, the search for a_{iz} can be done using a pre-selected set of values $\{a_j, j = 1, \ldots, m\}$ between the possible minimum objective measurement and the maximum objective measurement based on the current medical knowledge for the entire target population (such as normal range of human hemoglobin count). The pre-selected value a_j is not meant to be random, but rather fixed and ideally pre-determined before the realization of X_{ik}. For example, the normal range of human blood hemoglobin concentration can be determined from 5g/dL to 20g/dL so that a_{iz} is believed to be included in the range for any subject; if the increasing step is 1g/dL between a_j and a_{j+1}, then number of searching points, m, is equal to 16 in this case. The magnitude of the increasing step is determined by how precise the a_{iz} is expected to be. Again, this searching set is not considered random because it doesn't change with study or subject and may not be changed for decades, such as the normal range of human blood pressures.

Table 1 shows a toy example of how to estimate Q_{iz}. Note here, the number of searching points m need not necessarily be equal to the number (t) of clinical visits although we do so for illustration purpose. At each a_j, the outcome x_{ik} ($k = 1, \ldots, t$) is compared to a_j one at a time. Then the number of potential true positive (*TP*) and the number of potential true negative (*TN*) responses can be summarized per **Table 2**. For example, in the 1st data row of **Table 1** there are 9 $x_i \geq 5.0$ (positive) and only 6 g_i equal to one (PRO positive), therefore the *TP* is

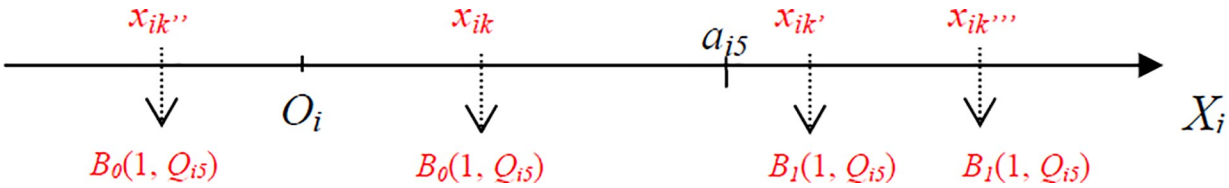

Fig 2. The g_{ik} is from two Bernoulli random variables with same parameter but opposite meaning depending where the X_{ik} is realized: $x_{ik} < a_{iz}$ or $\geq a_{iz}$.

https://doi.org/10.1371/journal.pone.0205845.g002

Table 1. Estimate of Q_{iz} based on 9 pairs of repeated outcomes (x_{ij}, g_{ij}) from subject *i*.

Samples	a_j	TP	TN	FP	FN	*TP+TN*	R_{ij}
(5, 0), (**7, 0**), (9, 0), (11, 1), (12, 1), (13, 1), (14, 1), (15, 1), (16, 1)	5.0	6	0	3	0	6	0.67
	6.0	6	1	2	0	7	0.78
	8.0	6	2	1	0	8	0.89
	10.0	**6**	**3**	**0**	**0**	**9**	**1.00**
	11.0	**6**	**3**	**0**	**0**	**9**	**1.00**
	12.0	5	3	0	1	8	0.89
	13.5	3	3	0	3	6	0.67
	15.0	2	3	0	4	5	0.56
	16.0	1	3	0	5	4	0.44
(5, 0), (**7, 1**), (9, 0), (11, 1), (12, 1), (13, 1), (14, 1), (15, 1), (16, 1)	5.0	7	0	2	0	7	0.78
	6.0	**7**	**1**	**1**	**0**	**8**	**0.89**
	8.0	6	1	1	1	7	0.78
	10.0	**6**	**2**	**0**	**1**	**8**	**0.89**
	11.0	**6**	**2**	**0**	**1**	**8**	**0.89**
	12.0	5	2	0	2	7	0.78
	13.5	3	2	0	4	5	0.56
	15.0	2	2	0	5	4	0.44
	16.0	1	2	0	6	3	0.33

Note: First sample shows: $\hat{Q}_{iz} = 1.00$, and the corresponding estimate of $a_{iz} = 10.5$

Second sample shows: $\hat{Q}_{iz} = 0.89$, and the corresponding estimate of $a_{iz} = 10$

https://doi.org/10.1371/journal.pone.0205845.t001

equal to 6 (see next paragraph for more details). The total number of such 2-by-2 tables is equal to m, as the total number of distinct a_j is m. The derivation in next subsection shows that the maximum of $R_{ij} = (TP+TN)_{ij}/t$ is a consistent estimator of Q_{iz}.

Table 1 shows how to use the pre-determined set of a_j ($j = 1, \ldots, m$) to calculate R_{ij} at each a_j based on two sets of 9 pairs of observations $(x_{i1}, g_{i1}) \ldots (x_{i9}, g_{i9})$ from Subject *i*. The only difference between the two sets of samples is the different values in the 2nd binary outcome g_{i2} (0 vs. 1). If the PRO_i is positive, $g_{ik} = 1$; otherwise $g_{ik} = 0$. The pre-determined set of a_j ($j = 1, \ldots, 9$) is listed in the 2nd column of Table 1. At each a_j, one can compare the 9 objective outcomes $(x_{i1}, \ldots, x_{i9})$ to a_j one at a time, and obtain the numbers of potential TP, FP, TN, FN per **Table 2** above. Thus, each data row of Table 1 displays the four statistics TP, FN, FP, and TN, corresponding to a_j. The estimate of Q_{iz} for Subject *i* at the $PROM_z$ is the maximum of R_{ij}. In this paper, if there are multiple tied maximums of R_{ij} the median of the corresponding a_j is used as an estimate of a_{iz}. This is because at each maximum of R_{ij}, the corresponding a_j could be an estimate of a_{iz}.

2.5 Derivation of the estimator of Q_{iz}

As illustrated in Fig 1 above, the Q_{iz} doesn't change its magnitude as long as $x_i \geq a_{iz}$ or $x_i < a_{iz}$ although Q_{iz} changes its meaning from conditional true positive rate (when $x_i \geq a_{iz}$) to

Table 2. Number of cell count at a_j ($j = 1, \ldots, m$) for subject *i* and $PROM_z$.

Objective efficacy outcome	Dichotomized PRO at $PROM_z$	
	PRO Positive	PRO Negative
$\geq a_j$	# of Potential Ture Positive (TP)	# of Potential False Positive (FP)
$< a_j$	# of Potential False Negative (FN)	# of Potential Ture Negative (TN)

https://doi.org/10.1371/journal.pone.0205845.t002

conditional true negative rate (when $x_i < a_{iz}$). This is a reasonable setup because the event of $PROM_i \geq z$ is a composite event including PRO_{iz}, PRO_{iz+1}, etc. For example, the event $PROM_i \geq 5$ includes $PRO_i = 5$, 6, or 7. When x_i is far above a_{iz}, Subject i may just give a higher PRO_i (say 7) and this event counts as one event of $PROM_i \geq 5$. This illustrates the fact that Q_{iz} can be independent of the distance between x_i and a_{iz}. Because we search a_{iz} such that $Pr(PROM_i \geq z \mid X_i \geq a_{iz}) = Pr(PROM_i < z \mid X_i < a_{iz})$ and Q_{iz} doesn't change its magnitude as long as $x_i \geq a_{iz}$ or $x_i < a_{iz}$., **we define $Q_{iz} = Pr(PROM_i \geq z \mid X_i \geq a, \forall\, a \geq a_{iz}) = Pr(PROM_i < z \mid X_i < b, \forall b < a_{iz})$** (see Fig 2 for the illustration), **where a and b are two arbitrary values in the objective measurement. Note here, although the clinical meaning of Q_{iz} changes from conditional positive rate to conditional negative rate according to $x_i \geq a_{iz}$ or $x_i < a_{iz}$, the magnitude of Q_{iz} doesn't change. This implies that the magnitude of Q_{iz} doesn't change with any subset of $X_i \geq a_{iz}$ or $X_i < a_{iz}$. In order to reflect the setup and the meaning of Q_{iz}, we use a and b here to indicate that Q_{iz} does not change its magnitude with any subset in $X_i \geq a_{iz}$ or $X_i < a_{iz}$.**

Also, the derivation of the Q_{iz} estimator doesn't assume independence among $X_{i1}, \ldots, X_{it}$. The cumulative distribution function of X_{i1} is denoted as F_{i1}. Because the x_{i1} is obtained in the 1st clinical visit before the realization of $X_{i2}, \ldots, X_{it}$, the cumulative distribution function of X_{ik} (denoted as F_{ik}, $k>1$) is the marginal cumulative distribution function, which can be obtained by integrating out $X_{i1}, \ldots, X_{ik-1}$ from the joint distribution $F_{Xi1, \ldots, Xik}$ for Subject i. The use of general form of F_{ik} in the derivation takes into consideration the correlated samples. The joint distribution $F_{Xi1, \ldots, Xik}$ applies to random variables with or without correlation. Therefore, the X_{ik} ($k = 1, \ldots, t$) are not assumed independent to each other and each X_{ik} has a different marginal distribution.

The derivation of the estimator of Q_{iz} starts with the probability of getting TN and TP at Visit k, which are presented in Expressions (1)–(4) below:

- When $a_j < a_{iz}$:

$$\begin{aligned} Pr_{ijk}(TN) &= Pr(X_{ik} < a_j \text{ and } G_{ik} = 0) \\ &= Pr(X_{ik} < a_j) Pr(G_{ik} = 0 | X_{ik} < a_j) \\ &= F_{ik}(a_j) Pr(G_{ik} = 0 | X_{ik} < a_j) \\ &= F_{ik}(a_j) Pr(G_{ik} = 0 | X_{ik} < a_j < a_{iz}) \\ &= F_{ik}(a_j) Q_{iz} \end{aligned} \tag{1}$$

With same argument, one can have the following:

$$\begin{aligned} Pr_{ijk}(TP) &= Pr(a_j < X_{ik} < a_{iz} \text{ and } G_{ik} = 1) + Pr(a_{iz} < X_{ik} \text{ and } G_{ik} = 1) \\ &= [F_{ik}(a_{iz}) - F_{ik}(a_j)] \bar{Q}_{iz} + \bar{F}_{ik}(a_{iz}) Q_{iz} \end{aligned} \tag{2}$$

- When $a_j \geq a_{iz}$:

$$\begin{aligned} Pr_{ijk}(TN) &= Pr(X_{ik} < a_{iz} \text{ and } G_{ik} = 0) + Pr(a_{iz} < X_{ik} < a_j \text{ and } G_{ik} = 0) \\ &= F_{ik}(a_{iz}) Q_{iz} + [F_{ik}(a_j) - F_{ik}(a_{iz})] \bar{Q}_{iz} \end{aligned} \tag{3}$$

$$Pr_{ijk}(TP) = Pr(a_j < X_{ik} \text{ and } G_{ik} = 1) = \bar{F}_{ik}(a_j) Q_{iz} \tag{4}$$

, where $\bar{Q}_{iz} = 1 - Q_{iz}$, $\bar{F}_{ik} = 1 - F_{ik}$.

Consequently, the expectation of TP+TN can be shown in Expressions (5) and (6), where E is the expectation operator.

- When $a_j \leq a_{iz}$:

$$E_{ij}(TP + TN) = \sum_{k=1}^{t}\{[F_{ik}(a_j) + \bar{F}_{ik}(a_{iz})]Q_{iz} + [F_{ik}(a_{iz}) - F_{ik}(a_j)]\bar{Q}_{iz}\} \quad (5)$$

- When $a_j > a_{iz}$:

$$E_{ij}(TP + TN) = \sum_{k=1}^{t}\{[\bar{F}_{ik}(a_j) + F_{ik}(a_{iz})]Q_{iz} + [F_{ik}(a_j) - F_{ik}(a_{iz})]\bar{Q}_{iz}\} \quad (6)$$

If a_j is equal to a_{iz}, both expressions (5) and (6) are reduced to tQ_{iz}. Therefore, R_{ij} = *(TP +TN)/t* is an unbiased estimator of Q_{iz} **only if** $a_j = a_{iz}$, and TP+TN follows the binomial distribution when $a_j = a_{iz}$ because its expectation follows the expectation of the binomial random variable (tQ_{iz}). We further notice that E_{ij} *(TP+TN)* obtains its maximum at a_{iz} when $Q_{iz} > 0.5$ (i.e. $Q_{iz} - \bar{Q}_{iz} > 0$) based on the sign of the derivative of E_{ij} *(TP+TN)* with respect to a_j. When $Q_{iz} > 0.5$, the derivative of E_{ij} *(TP+TN)* is positive at the left of a_{iz} (see Expression 5), and becomes negative at the right of a_{iz} (see Expression 6). Therefore, E_{ij} *(TP+TN)* not only reaches its maximum at a_{iz}, but also becomes tQ_{iz}. This is why the unbiased estimate of Q_{iz} is chosen as the maximum of R_{ij}. Similarly, E_{ij} *(TP+TN)* obtains its minimum at a_{iz} when $Q_{iz} < 0.5$.

In practice, it is reasonable to assume that a PROM has a non-negative association with the objective endpoint because it is obvious to see a potential direction of the PROM. If a negative association is expected, one can transform the objective outcome in order to have a non-negative association. For the example of a negative associate, if the PROM is the price satisfaction survey and the continuous objective endpoint is the cost of medical expense; then one can transform the cost by multiplying "-1" so that the higher negative cost (smaller cost) is in positive direction. Therefore, Q_{iz} can be assumed to be ≥ 0.5. If $Q_{iz} = 0.5$ (pure chance), this indicates that the PROM_z may not be able to reveal the truth; consequently, there is no conditional association between the *PROM* and the objective measurement X at PROM_z. This is because Q_{iz} is defined as the probability revealing truth at PROM_z; $Q_{iz} = 0.5$ is equivalent to Subject i flipping a fair coin to get the PRO_z by pure chance.

As discussed above, based on Expressions (5) and (6), the unbiased estimator of Q_{iz} is $\widehat{Q}_{iz} = max\{R_{ij} j = 1, \ldots, m\}$ if a_{iz} is in the searching set $\{a_j, j = 1, \ldots, m\}$. In practice, many tied maximums of R_{ij} may occur especially when t is small and m is large. In this case, the median of the tied maximums will be taken as the estimate. Because of this, $\widehat{Q}_{iz}$ becomes a consistent estimator. The variance of $\widehat{Q}_{iz}$ is nuisance because the validation of PROM is usually drawn from multiple subjects instead of Subject i. Nonetheless the variance estimate ($\widehat{Var}(\widehat{Q}_{iz})$) of $\widehat{Q}_{iz}$ for Subject i can be obtained by $\frac{\widehat{Q}_{iz}(1-\widehat{Q}_{iz})}{t}$, because TP+TN follows a binomial distribution with parameter t and Q_{iz} when $a_j = a_{iz}$. Further, because t is usually small, the exact binomial confidence interval for Q_{iz} is used for $\widehat{Q}_{iz}$ in the simulation study.

It is necessary to point out that if the two probabilities (say $Q_{\_iz}$ for negative truth and Q_{+iz} for positive truth) are not equal, many more assumptions are needed to estimate $Q_{\_iz}$ and Q_{+iz}. Using our method, when both $Q_{\_iz}$ and Q_{+iz} are both greater than 0.5 we can have $t[rF_{ik}(a_{iz}) + \bar{F}_{ik}(a_{iz})]Q_{+iz} = \underset{a_j}{Max}(TP + TN)$, where $r = Q_{+iz}/Q_{\_iz}$. We can estimate a_{iz}using a_j at which the maximum of $(TP + TN)$ is reached, but we have unknown r and many unknown F_{ik} ($k = 1, \ldots, t$). If we further assume r is known, we still could not find the estimate for Q_{+iz}

because we don't know these F_{ik}. Unless we further assume the distribution function of X_{ik} at each clinical visit k, we can have a consistent estimate of $Q_{\_iz}$ and Q_{+iz}. But we feel that these further assumptions on knowing r and F_{ik} ($k = 1, \ldots, t$) are not practical, especially in medical device clinical trials. Therefore, we only search the threshold such that the two probabilities are equal in this paper.

2.6 Inference of Q_{iz} in multiple subjects

So far, Q_{iz} is estimated based on t repeated pairs of measurements from Subject i for the PROM_z. If one wants to know the population Q_z for the *PROM* and the objective measurement X at PROM_z in a target patient population, the Q_z can be confirmed by the mean $(\sum_{i=1}^{n} \widehat{Q}_{iz}/n)$ of $\widehat{Q}_{iz}$ with its 95% CI. For example, the lower bound of the 95% confidence interval of Q_z must be greater than a desired probability of revealing truth in order for one to believe that the PROM_z is able to reveal disease status for majority of subjects in the patient population.

The ability of the $PROM_i$ to detect a change in the objective endpoint X_i could be confirmed by the statistically significant change of a_{iz} to $a_{iz'}$ obtained by different dichotomizations of the PRO_i. Note, the magnitude of a_{iz} will be changed when the PRO_i is dichotomized differently. For example, the PRO_i can be dichotomized at scale 7 by "at least very much improvement or otherwise" or at scale 6 by "at least much improvement or otherwise". This change of dichotomization represents one unit change of the PRO_i from scale 6 to 7, and thus the change of a_{i6} to a_{i7} measures the ability of the $PROM_i$ to detect the change in the objective endpoint X_i. The a_{iz} is expected to be larger when the PRO_i is dichotomized by "at least very much improvement or otherwise" compared to that by "at least much improvement or otherwise". This is because "at least very much improvement" is more difficult to be reached and thus its minimum threshold is expected to be higher than that for "at least much improvement". One can obtain the estimate of the change of a_{iz} to $a_{iz'}$ from each of n different subjects, and perform the test of the mean change > 0.

3 Simulation and illustration

3.1 Simulation

Simulation data from Subject i is used to illustrate the characteristics of $\widehat{Q}_{iz}$, especially to show $\widehat{Q}_{iz}$ is a consistent estimator of Q_{iz}. The simulation is not meant to align with a real clinical trial, however the use of $\widehat{Q}_{iz}$ in a clinical trial is presented after the simulation using hypothetical clinical data. Because Q_{iz} is defined at subject level, the simulation uses one treatment for a disease in one subject only. The primary endpoint is an objective endpoint measuring the change of the disease status from baseline to 3 months. The PROM is the 7-scaled disease-related satisfaction PROM such as illustrated in Fig 1. In order to include different means and standard deviations, the simulation uses 5 different means [$\boldsymbol{\mu} = (0, 0.5, 1, 1.5, 2)$] and 5 associated different standard deviations [$\boldsymbol{\sigma^2} = (1, 1.3, 1.6, 1.9, 2.2)$] as two building blocks to construct various multivariate normal distribution for the objective endpoint. For example, if $t = 10$ then X_{ik} ($k = 1, \ldots 10$) will follow the multivariate normal distribution with stacked mean vector $(\boldsymbol{\mu}, \boldsymbol{\mu})$ and the variance-covariance matrix with diagonal elements of $\boldsymbol{\sigma^2}$ repeated similarly on diagonal and the off-diagonal element of $\rho\sigma_t\sigma_s$. Other setups are described as follows:

a. The correlation coefficient (ρ) between X_{ik} and $X_{ik'}$ ranges from 0.3, 0.5, and 0.8.

b. 'a_{i7}' (the minimum objective threshold for "at least very much improved") is equal to 1.2, 'a_{i3}' is equal to -0.3 and 'a_{i5}' is equal to 0.4.

c. The underlying probability of revealing the truth, Q_{iz} (z = 3, 5, or 7) has values of 0.5, 0.6, 0.7, 0.8, and 0.9.

d. Number of repeated measurements for the subject is t = 5, 10, 20, 40.

e. Pre-selected a_j ranges from -2 to 5.0 with increasing step of 0.1, therefore m = 71. Because the minimum two standard deviations below the five means is -2 and the maximum two standard deviations above the five means is 5, this range is wide enough to include all underlying true values of a_{i3}, a_{i5}, and a_{i7}.

f. Number of simulation is 10,000.

For each combination of ρ (0.3, 0.5, 0.8), a_{iz} (-0.3, 0.4, 1.2), and Q_{iz} (0.5, 0.6, 0.7, 0.8, 0.9), the t (5, 10, 20, 40) pairs of outcomes (x_{ik}, g_{ik}) (k = 1, . . ., t) are sampled as follow. First x_{ik} (k = 1, . . ., t) is drawn from the corresponding multivariate normal distribution. If $x_{ik} \geq a_{iz}$, g_{ik} is drawn from *Bernoulli* (Q_{iz}); otherwise g_{ik} is drawn from *Bernoulli* (1-Q_{iz}). Then an estimate of Q_{iz} is calculated using the method described above based on the t pairs of outcomes, and its 95% CI is calculated using the exact binomial confidence interval due to small samples. These steps are repeated 10,000 times for each underling value of Q_{iz} and t; and then the mean of these 10,000 $\widehat{Q}_{iz}$ and the coverage probability of the 95% CIs for the Q_{iz} are obtained.

Figs 3–5 show three examples that the mean of these 10,000 $\widehat{Q}_{iz}$ converges to Q_{iz} regardless of the values of ρ and a_{iz}. As the number of clinical visits increases for Subject i, the mean of $\widehat{Q}_{iz}$ approaches its underlying true value of Q_{iz}. The converging pattern exists for every value of Q_{iz} (0.6, 0.7, 0.8, 0.9) except for Q_{iz} = 0.5. This is not a surprise because when Q_{iz} = 0.5 there is no association between $PROM_i$ and X_i at $PROM_z$. As shown in expressions (5) and (6), when Q_{iz} = 0.5 every R_{ij} (j = 1 . . . m) is an unbiased estimator of Q_{iz}. A separate simulation using the median of R_{ij} as $\widehat{Q}_{iz}$ is performed when Q_{iz} = 0.5. The mean $\widehat{Q}_{iz}$ ranges from 0.50 to 0.52 (converging to 0.5) for different combinations of ρ, a_{iz}, Q_{iz}, and t. In practice, the simulation results for Q_{iz} = 0.5 in Figs 3–5 can be used as a reference to set a minimum acceptable Q_{iz} value. Table 3 shows that mean $\widehat{Q}_{i7}$ is a fairly close estimate of Q_{i7} under different values of t (5, 10, 20, 40). It is found that the probability of the 95% CI including the true value of Q_{i7} (coverage probability) is at least 95% due to the use of exact binomial confidence interval.

3.2 Case study: Hypothetical clinical trial data

The probability Q_{iz} of revealing truth for Subject i at $PROM_z$, has been applied to hypothetical clinical trial data in order to assess the conditional association parameter in multiple subjects. The purpose of the trial is to improve near vision by a medical device. Each subject had a test device implanted and was followed up at Months 3, 6, 12, 18, 24, 30 post procedure. At each follow-up visit, a subject had his/her uncorrected near visual acuity (UCNVA) measured using ETDRS Chart at 40 cm/16 in, and answered a unidimensional PROM question with 7 possible outcomes as shown in Fig 1. The question in the PROM was "*How satisfied are you with your near vision without reading glasses after the treatment*?" The change from baseline in UCNVA is considered as the continuous objective clinical endpoint with a larger change indicating better near vision. The outcome of the satisfaction question is the *PRO* which can be dichotomized in 3 ways for every subject: $\geq$5 or otherwise, $\geq$6 or otherwise, $\geq$7 or otherwise. The mean $\widehat{Q}_{iz}$ (z = 5, 6, or 7) is used to assess the probability of the $PROM_z$ to reveal the status of the visual acuity in the targeted population.

The pre-determined threshold searching set {a_j, j = 1, . . ., m} ranges from -20 to 60 letters with an increasing step of 1. This set contains m = 81 searching points for the minimum

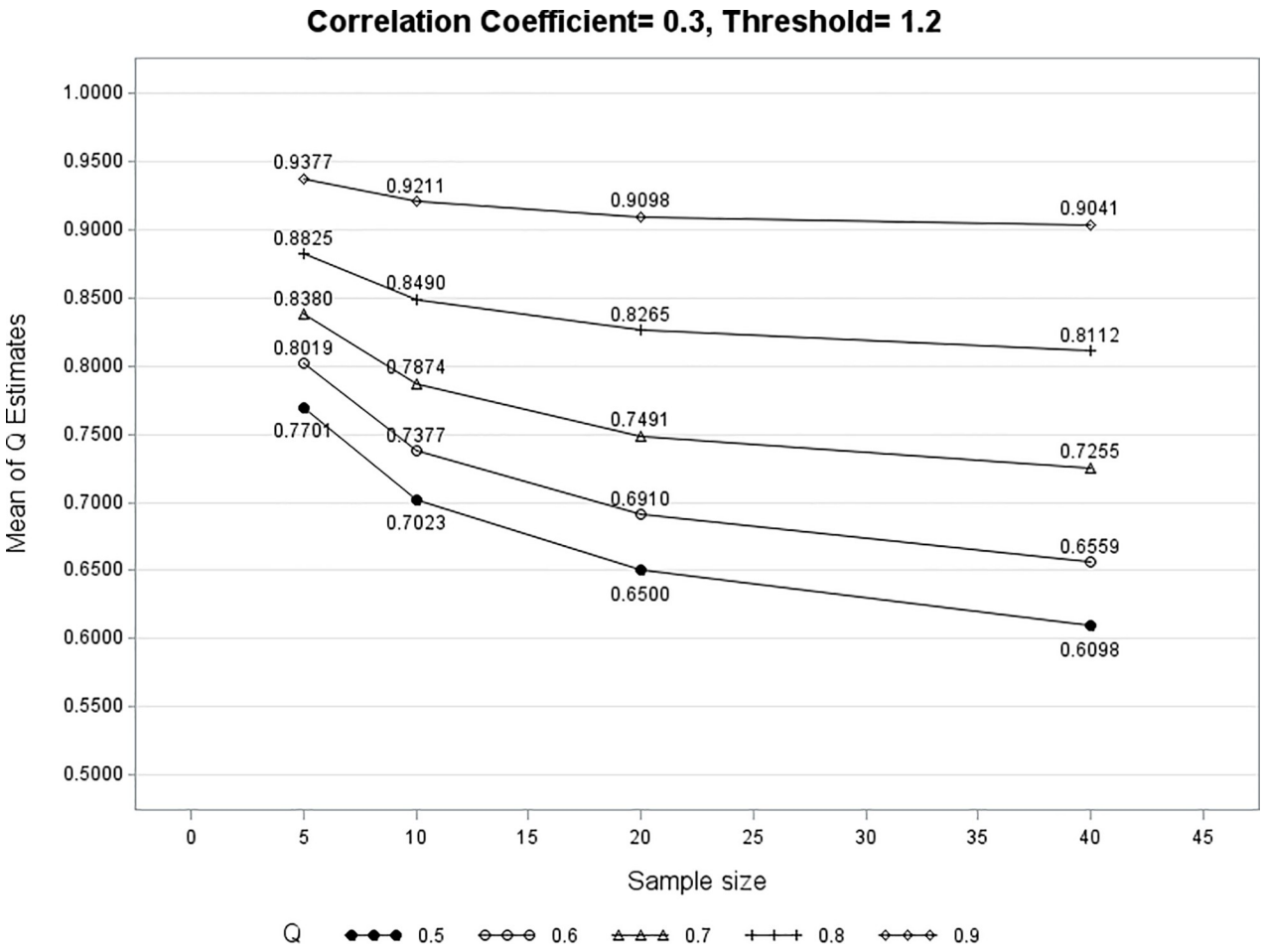

Fig 3. The mean $\widehat{Q}_{iz}$ converges to its underlying value of Q_{iz} as sample size increases ($\rho = 0.3$, $a_{iz} = 1.2$).

https://doi.org/10.1371/journal.pone.0205845.g003

threshold a_{iz} ($z = 5, 6$, or 7). It is believed that the threshold-searching set is large enough to contain the true value of a_{iz} for PROM_z for every subject in the target population.

Table 4 below shows that the mean of the $\widehat{Q}_{iz}$ (probability of revealing truth) and the mean $\widehat{a}_{iz}$ in the change of UCNVA. As expected, one can see that the highest satisfaction has the lowest mean probability of revealing truth uncorrected visual acuity and the largest threshold in the change of UCNVA: 21 more letters correctly identified from baseline. The associated 95% CIs for Q_{iz} well exclude 0.5 indicating Q_{iz} from the majority of subjects are greater than 0.5 and consequently the probability of the PROM_z revealing subjects' uncorrected visual acuity is established. Since the PROM_z has $> 83\%$ probability (based on the lower limits) of revealing the status of UCNVA, it may be used as a binary endpoint for the primary inference for uncorrected near visual acuity.

Table 5 shows the median of $\widehat{a}_{iz} - \widehat{a}_{iz'}$ when the satisfaction level changes. The $\widehat{a}_{iz} - \widehat{a}_{iz'}$ is found to have a highly skewed distribution; therefore p-values are reported here from a non-parametric signed rank test, and the reference statistic is referred to median instead of mean. One can observe that

1. When the *PRO* increases from ≥ 5 to ≥ 6, the majority of subjects have no change (median = 0) in their uncorrected near vision acuity; this means that the *PRO* change from scale 5 to scale 6 may not represent a change in majority subjects' uncorrected near vision acuity.

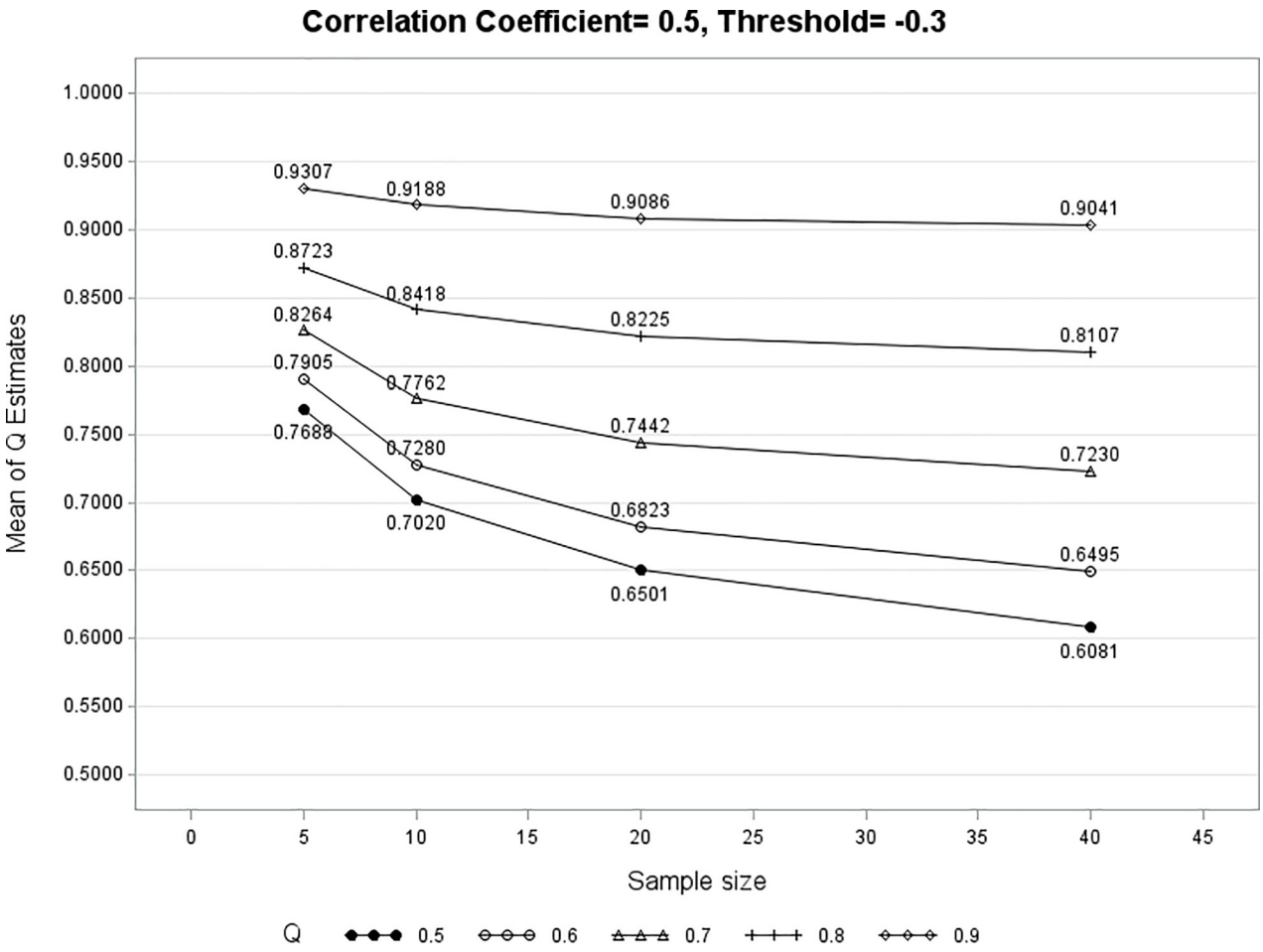

Fig 4. The mean $\widehat{Q}_{iz}$ converges to its underlying value of Q_{iz} as sample size increases (ρ = 0.5, a_{iz} = -0.3).

https://doi.org/10.1371/journal.pone.0205845.g004

2. When the *PRO* increases from $\geq$6 to $\geq$7 or $\geq$5 to $\geq$7, the majority of subjects have a positive change (median = 9 or 21, respectively) in their uncorrected near vision acuity; this means that the *PRO* changes from a lower score to 7 represent a change in majority subjects' uncorrected near vision acuity.

These indicate that a change of one PROM unit in this case might not be adequate for a translation to a change in uncorrected near visual acuity. An increase of at least two (2) PROM units represents that the majority subjects have a positive increase in their uncorrected near visual acuity. Consequently, the ability of detecting the change of uncorrected near vision function by this PROM is suggested by two (2) PROM units in this clinical trial instead of one (1) PROM unit; or the majority of subjects have their PRO scores changed to 7. It is noted that the number of samples from each subject is $\leq$ 6 in this trial, which limits the capability of this method to search for a_{iz}.

4 Concluding remarks

The conditional probability Q_{iz} revealing the true status of Subject *i*'s disease at PROM_z is a new quantitative statistic assessing the conditional association between a unidimensional $PROM_i$ and a continuous objective endpoint X_i measuring the disease status. The probability

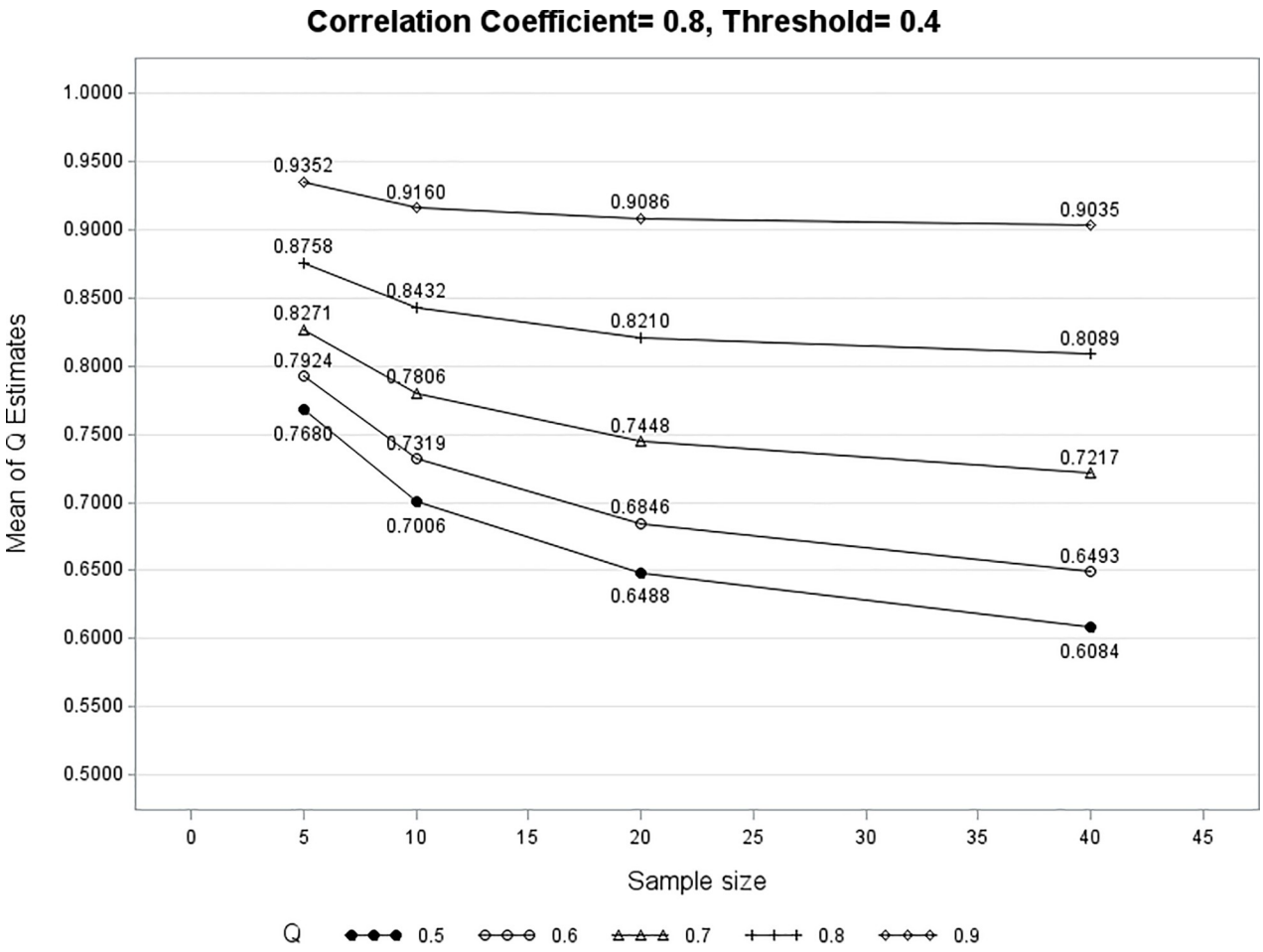

Fig 5. The mean $\widehat{Q}_{iz}$ converges to its underlying value of Qiz as sample size increases (ρ = 0.8, aiz = 0.4).

https://doi.org/10.1371/journal.pone.0205845.g005

Q_{iz} of revealing truth is estimated for each subject using paired observations (x_{ik}, g_{ik}) measured repeatedly at different clinical visits (such as Months 3, 6, 12 etc.). The Q_{iz} reveals truth with respect to the latent minimum objective threshold a_{iz} (i.e. $x_{ik} \geq a_{iz}$, or $x_{ik} < a_{iz}$). When the $PROM_i$ doesn't associate with the objective endpoint X_i, the Q_{iz} is equal to the pure chance of 0.5. Because Q_{iz} is a probability measure, this situation looks like one has flipped a fair coin to

Table 3. Mean estimate and coverage probability of Q_{i7} ρ = 0.8, a_{i7} = 1.2.

		True value of Q				
		0.5	**0.6**	**0.7**	**0.8**	**0.9**
t = 40	$\widehat{Q}$	0.61	0.65	0.72	0.81	0.90
	Coverage Probability of the 95% CI	0.851	0.958	0.961	0.981	0.971
t = 20	$\widehat{Q}$	0.65	0.69	0.75	0.82	0.91
	Coverage Probability of the 95% CI	0.841	0.949	0.986	0.982	0.994
t = 10	$\widehat{Q}$	0.70	0.73	0.78	0.84	0.92
	Coverage Probability of the 95% CI	0.915	0.980	1.000	1.000	0.996
t = 5	$\widehat{Q}$	0.77	0.79	0.83	0.88	0.94
	Coverage Probability of the 95% CI	1.000	1.000	1.000	1.000	1.000

https://doi.org/10.1371/journal.pone.0205845.t003

Table 4. Mean $\widehat{Q}_{iz}$ and mean $\widehat{a}_{iz}$ in the change of UCNVA.

Satisfaction dichotomized value	# of subjects *	Mean $\widehat{Q}_{iz}$ (97.5% CI)	Mean $\widehat{a}_{iz}$ (# letters correctly identified)
≥5	414	0.91 (0.893, 0.917)	8
≥6	324	0.88 (0.864, 0.893)	14
≥7	190	0.85 (0.831, 0.868)	21

* includes subjects whose PROs contain the dichotomized value and have at least two different objective outcomes

https://doi.org/10.1371/journal.pone.0205845.t004

get his/her *PRO* regardless the status of his/her disease. When a PROM is used as a measure for a disease/condition in a clinical trial setup, the probability of revealing truth must be at least statistically greater than the pure chance of 0.5.

The threshold searching set $\{a_j: j = 1, \ldots, m\}$ can be pre-determined using the current clinical standard of the possible minimum and maximum objective measurements in the target population. For example, the human hemoglobin concentration ranges from 5 g/dL to 20 g/dL. The number m can be determined based on how precise a_{iz} is expected to be.

In practice, a clinical trial has n subjects and thus has n estimates of Q_{iz} $(i = 1, \ldots, n)$. In order to have the PROM_z used for a target population, the majority of Q_{iz} $(i = 1, \ldots, n)$ have to be greater than the pure chance of 0.5; or it is equivalent to say that the mean/median of the Q_{iz} $(i = 1, \ldots, n)$ should be greater than 0.5. Although the mean/median of the $Q_{iz} > 0.5$ would indicate some association between the *PROM* and the objective endpoint X greater than chance in the target population, a higher quality *PROM* should have a larger value of the mean/median of the Q_{iz} $(i = 1, \ldots, n)$. Let's denote δ as the minimum value of the mean/median of the Q_{iz} $(i = 1, \ldots, n)$ which is an acceptable probability for PROM_z to reveal the status of the majority of subjects' disease. To confirm that the majority of subjects have their Q_{iz} $(i = 1, \ldots, n)$ greater than δ, one can simply test that the mean/median of the Q_{iz} $(i = 1, \ldots, n)$ among n different subjects is $>\delta$.

When the *PRO* is dichotomized differently by one PROM unit increased at a time, one can get the associated estimate of the change of the minimum threshold in the objective measurement for each subject, such as $\widehat{a}_{iz} - \widehat{a}_{iz'}$ $(i = 1, \ldots, n)$. If the mean of these estimates from different subjects is statistically significantly greater than 0, then the PROM has the ability to detect a change in the objective endpoint. In case that $\widehat{a}_{iz} - \widehat{a}_{iz'}$ $(i = 1, \ldots, n)$ has a skewed distribution, one should use the median of the estimates of $\widehat{a}_{iz} - \widehat{a}_{iz'}$ $(i = 1, \ldots, n)$ so that the test implies that the majority of $a_{iz} - a_{iz'}$ $(i = 1, \ldots, n)$ are greater than 0.

The limitations of using Q_{iz} include (1) it is applicable to a unidimensional PROM or a PROM item of interest in a multi-dimensional PROM instrument when a valid continuous objective measure of the disease status exists, and (2) if the number of repeated measurements is small, the estimator of Q_{iz} is more biased. In this case, one can adjust the minimum

Table 5. Ability of detecting a change: Median of $\widehat{a}_{iz} - \widehat{a}_{iz'}$.

Satisfaction Change	# of subjects *	change of $\widehat{a}_{iz} - \widehat{a}_{iz'}$		p-value by Signed Rank Test
		Mean	Median	
From ≥5 to ≥6	324	11	0	<0.001
From ≥6 to = 7	190	15	9	<0.001
From ≥5 to = 7	190	20	21	<0.001

*includes all subjects who are in Table 3 and have a change objective value when the associated PRO changes

https://doi.org/10.1371/journal.pone.0205845.t005

acceptable probability of revealing truth in order to have confidence for the $PROM_z$ to reveal truth. Further research may focus on a quantitative method for measuring the conditional association between a multi-dimensional PROM and a pertinent objective measurement.

Appendix A: Notations

- Sub-indexes i and j represent Subject i and threshold searching point j within a clinical visit k ($i = 1, \ldots, n$, $j = 1, \ldots, m$, and $k = 1, \ldots, t$). The letter z denotes the z^{th} scale of the PROM ($PROM_z$).
- The a_{iz} is a fixed parameter which is defined as the minimum latent threshold in terms of the objective measurement for Subject i at $PROM_z$. The a_{iz} is defined for the z^{th} scale and Subject i. For example, if the PROM has 5 different scales, then we will have five different values of a_{iz} for the subject.
- The a_j is the j^{th} searching point for a_{iz}, and the a_j belongs to a fixed pre-selected threshold searching set $\{a_j: j = 1, \ldots, m\}$ (such as the normal range of hemoglobin count with an increasing step of 0.5). The a_j is a nonrandom variable and does not change with subject. The set is selected based on the current clinical standard of normal range.
- The X is the random variable for the continuous objective measurement of the status of a subject's disease/condition, and lower case x is an outcome/realization of X.
- G_{ik}^1 is the Bernoulli random variable with probability Q_{iz} to be 1 when $x_{ik} \geq a_{iz}$.
- G_{ik}^0 is the Bernoulli random variable also with probability Q_{iz} to be 0 when $x_{ik} < a_{iz}$.
- G_{ik} represents two mixed Bernoulli random variables with the same parameter Q_{iz} (but opposite meaning) G_{ik}^1 (if $x_{ik} \geq a_{iz}$) or G_{ik}^0 (if $x_{ik} < a_{iz}$).

Acknowledgments

The authors acknowledged the administrative support of this research work from the Division of Biostatistics, Office of Surveillance and Biometrics, CDRH, FDA.

Author Contributions

Conceptualization: Chul Ahn, Xin Fang.

Data curation: Xin Fang.

Formal analysis: Xin Fang.

Investigation: Xin Fang.

Methodology: Chul Ahn, Xin Fang, Phyllis Silverman, Zhiwei Zhang.

Writing – original draft: Xin Fang.

Writing – review & editing: Chul Ahn, Xin Fang, Phyllis Silverman, Zhiwei Zhang.

References

1. Willke RJ, Burke LB, Erickson P. Measuring treatment impact: A review of patient reported outcomes and other efficacy variables in approved product labels. Controlled Clinical Trials 2004; 25: 535–552. https://doi.org/10.1016/j.cct.2004.09.003 PMID: 15588741

2. Gnanasakthy A., Mordin M, Clark M., DeMuro C., Fehnel S., Copley-Merriman C. A Review of Patient-Reported Outcome Labels in the United States: 2006 to 2010. Value in Health 2012;, 15(3): 437–442. https://doi.org/10.1016/j.jval.2011.11.032 PMID: 22583453
3. Patrick DL, Burke LB, Powers JH, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Health 2007; 10 Supp. 2: S125–137.
4. Ader DN. Developing the Patient-Reported Outcomes Measurement Information System (PROMIS). Medical Care 2007; 45(Suppl. 1): S1–S2.
5. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap Cooperative Group during its first two years. Medical Care 2007; 45(Suppl. 1): S3–S11.
6. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric Evaluation and Calibration of Health-Related Quality of Life Item Banks Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Medical Care 2007; 45: S22–S31. https://doi.org/10.1097/01.mlr.0000250483.85507.04 PMID: 17443115
7. Cappelleri JC, Bushmakin AG. Interpretation of patient reported outcomes. Statistical Methods in Medical Research 2014; 23(5): 460–483 https://doi.org/10.1177/0962280213476377 PMID: 23427226
8. Massof RW. A general theoretical framework for interpreting patient-reported outcomes estimated from ordinally scaled item responses. Statistical Methods in Medical Research 2014; 23(5) 409–429. https://doi.org/10.1177/0962280213476380 PMID: 23427227
9. FDA Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009. https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf
10. Fayers PM, Machin D. Quality of Life: The assessment, analysis and interpretation of patient-reported outcomes. 2007, 2nd, John Wiley & Sons Ltd.
11. Bassim CW, Fassil H, et al (16 other authors). Validation of the National Institutes of Health chronic GVHD Oral Mucosal Score using component-specific measures. Bone Marrow Transplantation 2014, 49: 116–121. https://doi.org/10.1038/bmt.2013.137 PMID: 23995099
12. Bredart A, et al (with other 7 authors). Quality of care in the oncology outpatient setting from patients' perspective: a systematic review of questionnaires' content and psychometric performance. Psycho-Oncology 2015, 24(4): 382–394. https://doi.org/10.1002/pon.3661 PMID: 25196048
13. Brod M, Hojbjerre L, Bushnell DM, Hansen CT. Assessing the impact of non-severe hypoglycemic events and treatment in adults: development of the Treatment-Related Impact Measure—Non-severe Hypoglycemic Events (TRIM-HYPO). Quality of Life Research 2015, 24: 2971–2984. https://doi.org/10.1007/s11136-015-1023-6 PMID: 26094008
14. Cavaletti G. et al (with 27 authors). The chemotherapy-induced peripheral neuropathy outcome measures standardization study: from consensus to the first validity and reliability findings. Annal of Oncology 2013, 24 (2): 454–462.
15. Cella D, Jensen SE, Webster K, Du H, Lai JS, Rosen S, Tallman MS, Yount S. Measuring Health-Related Quality of Life in Leukemia: The Functional Assessment of Cancer Therapy–Leukemia (FACT-Leu) Questionnaire. Value in Health 2012, 15(8): 1051–1058. https://doi.org/10.1016/j.jval.2012.08.2210 PMID: 23244807
16. Coyne KS, Margolis MK, Murphy J, Spies J. Validation of the UFS-QOL-Hysterectomy Questionnaire: Modifying an Existing Measure for Comparative Effectiveness Research. Value in Health 2012, 15(5): 674–679. https://doi.org/10.1016/j.jval.2012.03.1387 PMID: 22867776
17. Gewandtera JS, et al (with other 10 authors). Associations between a patient-reported outcome (PRO) measure of sarcopenia and falls, functional status, and physical performance in older patients with cancer. Journal of Geriatric Oncology 2015, 6(6): 433–441. https://doi.org/10.1016/j.jgo.2015.07.007 PMID: 26365897
18. Gunaydin G, et al (with other 7 authors). Cross-cultural adaptation, reliability and validity of the Turkish version of the Japanese Orthopaedic Association Back Pain Evaluation Questionnaire. Journal of Orthopaedic Science 2016, 21(3):295–8. https://doi.org/10.1016/j.jos.2016.01.006 PMID: 26898339
19. Harris-Hayes M, McDonough CM, Leunig M, Lee CB Callaghan JJ, Roos Em. *Clinical outcomes assessment in clinical trials to assess treatment of femoroacetabular impingement: use of patient-reported outcome measures*. Journal of the American Academy of Orthopaedic Surgeons 2013, 21 (01): S39–S46.
20. Kimball AB, et al (with other 9 authors). Assessing the validity, responsiveness and eaningfulness of the Hidradenitis Suppurativa Clinical Response (HiSCR) as the clinical endpoint for idradenitis suppurativa treatment. British Journal of Dermatology 2014, 171(6): 1434–1442. https://doi.org/10.1111/bjd.13270 PMID: 25040429

21. Kopjar B, Tetreault L, Kalsi-Ryan S, Fehlings M. Psychometric properties of the modified Japanese Orthopaedic Association scale in patients with cervical spondylotic myelopathy. Spine 2015, 40(1): E23–E28 https://doi.org/10.1097/BRS.0000000000000648 PMID: 25341993

22. Pinheiro LC, Callahan LF, Cleveland RJ, Edwards LJ, Reeve BB. The Performance and Association Between Patient-reported and Performance-based Measures of Physical Functioning in Research on Individuals with Arthritis. The Journal of Rheumatology 2016, 43(1): 131–137. https://doi.org/10.3899/jrheum.150432 PMID: 26628600

23. Motl RW, Cadavid D, Sandroff BM, Pilutti LA, Pula JH, Benedict RHB. Cognitive processing speed has minimal influence on the construct validity of Multiple Sclerosis Walking Scale-12 scores. Journal of the Neurological Sciences 2013, 335(1–2): 169–173. https://doi.org/10.1016/j.jns.2013.09.024 PMID: 24104065

24. Petrillo J, Cano SJ, McLeod LD, Coon CD. Using Classical Test Theory, Item Response Theory, and Rasch Measurement Theory to Evaluate Patient-Reported Outcome Measures: A Comparison of Worked Examples. Value in Health, Vol 2015, 18(1): 25–34 https://doi.org/10.1016/j.jval.2014.10.005 PMID: 25595231

25. Selman L, et al (with other 13 authors). 'Peace'and 'life worthwhile'as measures of spiritual well-being in African palliative care: a mixed-methods study. Health and Quality of Life Outcomes 2013, 11: 94. https://doi.org/10.1186/1477-7525-11-94 PMID: 23758738

26. Strober BE, Nyirady J, Mallya UG, Guettner A, Papavassilis C, Gottlieb A, Elewski BE, Turner-Bowker DM, et al. Item-Level Psychometric Properties for a New Patient-Reported Psoriasis Symptom Diary. Value in Health 2013, 16(6): 1014–1022 https://doi.org/10.1016/j.jval.2013.07.002 PMID: 24041351

27. Tubergen AV, Black P, McKenna SP, Coteur G. FRI0523 Validity of ankylosing spondylitis patient-reported outcome instruments in the broad axial spondyloarthritis population. Annals of the Rheumatic Diseases, 2013, 72: A551

28. Turk DC, Dworkin RH, Trudeau JJ, Benson C, Biondi DM, Katz NP, Kim M. Validation of the Hospital Anxiety and Depression Scale in Patients With Acute Low Back Pain 2015, 16(10): 1012–1021. https://doi.org/10.1016/j.jpain.2015.07.001 PMID: 26208762

29. Vascellari A, Schiavetti S, Rebuzzi E, Coletti N. Translation, cross-cultural adaptation and validation of the Italian version of the Nottingham Clavicle Score (NCS). Archives of Orthopaedic and Trauma Surgery 2015, 135:1561–1566 https://doi.org/10.1007/s00402-015-2301-z PMID: 26254581

30. Zeneli A, Fabbri E, Donati G, Tierney G, Pasa S, Berardi MA, Maltoni M. Translation of Supportive Care Needs Survey Short Form 34 (SCNS-SF34) into Italian and cultural validation study. Support Care Cancer 2016, 24:843–848. https://doi.org/10.1007/s00520-015-2852-0 PMID: 26166001

31. Rejas J, Ruiz M, Pardo A, Soto J. Detecting Changes in Patient Treatment Satisfaction with Medicines: The SATMED-Q. Value in Health 2013, 16(1): 88–96. https://doi.org/10.1016/j.jval.2012.08.2224 PMID: 23337219

32. Giesinger JM, Kuster MS, Behrend H, Giesinger K. Association of psychological status and patient-reported physical outcome measures in joint arthroplasty: a lack of divergent validity. Health and Quality of Life Outcomes 2013, 11: 64. https://doi.org/10.1186/1477-7525-11-64 PMID: 23601140

33. Manor L, et al (with other 6 authors). Age-related variables in childhood epilepsy: How do they relate to each other and to quality of life? Epilepsy & Behavior 2013, 26(1): 71–74.

34. Persson E, Eklund M, Lexell J, Rivano-Fischer M. Psychosocial coping profiles after pain rehabilitation: associations with occupational performance and patient characteristics. Disability and Rehabilitation 2016, https://doi.org/10.3109/09638288.2016.1141243 PMID: 26883399

35. Tate FR. Correlation between a Discrete and a Continuous Variable. Point-Biserial Correlation. The Annals of Mathematical Statistics 1954, 25(3): 603–607.

36. Rasch G. Probability models for some intelligence and attainment tests. Copenhagen, Danish Institute for Educational Research, 1960.

37. Kim M, Wall MM, Li G. Risk stratification for major postoperative complications in patients undergoing intra-abdominal general surgery using latent class analysis. Anesthesia & Analgesia 2018 Mar, 126 (3):848–857. https://doi.org/10.1213/ANE.0000000000002345 PMID: 28806210

38. Lombardi S, Santini G, Marchetti GM, Focardi S. Generalized structural equations improve sexual-selection analyses. PLOS ONE 2017, https://doi.org/10.1371/journal.pone.0181305

39. Lin L., Hedayat AS., Wu W. Statistical Tools for Measuring Agreement. Springer New York Dordrecht Heidelberg London, 2012.