Taylor & Francis
Taylor & Francis Group

Expert Review of Pharmacoeconomics & Outcomes Research

ISSN: 1473-7167 (Print) 1744-8379 (Online) Journal homepage: http://www.tandfonline.com/loi/ierp20

Statistical challenges of quality of life and cancer: new avenues for future research

Laurent Boyer, Karine Baumstarck, Pierre Michel, Mohamed Boucekine, Amelie Anota, Franck Bonnetain, Joel Coste, Bruno Falissard, Alice Guilleux, Jean-Benoit Hardouin, Anderson Loundou, Mariette Mercier, Mounir Mesbah, Alexandra Rouquette, Veronique Sebille, Mathilde GE Verdam, Badih Ghattas, Francis Guillemin & Pascal Auquier

To cite this article: Laurent Boyer, Karine Baumstarck, Pierre Michel, Mohamed Boucekine, Amelie Anota, Franck Bonnetain, Joel Coste, Bruno Falissard, Alice Guilleux, Jean-Benoit Hardouin, Anderson Loundou, Mariette Mercier, Mounir Mesbah, Alexandra Rouquette, Veronique Sebille, Mathilde GE Verdam, Badih Ghattas, Francis Guillemin & Pascal Auquier (2014) Statistical challenges of quality of life and cancer: new avenues for future research, Expert Review of Pharmacoeconomics & Outcomes Research, 14:1, 19-22

To link to this article: http://dx.doi.org/10.1586/14737167.2014.873704

EXPERT REVIEWS

Statistical challenges of quality of life and cancer: new avenues for future research

Expert Rev. Pharmacoecon. Outcomes Res. 14(1), 19–22 (2014)

Laurent Boyer*, Karine Baumstarck, Pierre Michel, Mohamed Boucekine, Amelie Anota, Franck Bonnetain, Joel Coste, Bruno Falissard, Alice Guilleux, Jean-Benoit Hardouin, Anderson Loundou, Mariette Mercier, Mounir Mesbah, Alexandra Rouquette, Veronique Sebille, Mathilde GE Verdam, Badih Ghattas, Francis Guillemin and Pascal Auquier

**Author for correspondence: laurent.boyer@ap-hm.fr*

For a full list of author affiliations, please see page 22.

Statistical modeling conference on the quality of life measurements of the French National Platform of Quality of Life and Cancer
Faculty of Science in Luminy, Marseille, France, 12–13 September 2013

The French National Platform of Quality of Life and Cancer and the statistical team of the Mathematical Institute of Luminy undertook a successful first conference addressing the statistical challenges of measuring the quality of life in the field of oncology. More than 15 presentations were made over a 2-day period by the Faculty of Sciences in Luminy. The conference managed to assemble participants from different disciplines, such as mathematics and statistics, public health, epidemiology and psychology, to debate the key statistical and methodological issues of quality of life measurement and analysis. Three main topics were covered in this conference: the treatment of missing data, the development of item banking and computerised adaptive testing and the detection/understanding of response shift.

Quality of life (QoL) assessments are becoming increasingly important to the evaluation of the treatments and care that are provided to patients with chronic diseases, especially in the field of oncology. However, QoL assessment remains underutilized in clinical practice and policy decision-making. The objective of this conference was to discuss new statistical challenges and perspectives in QoL research to enhance the use of QoL measures in clinical practice for patients with chronic diseases. In the inaugural presentation, B Falissard recalled the main advances in the metrological and statistical methods used to develop QoL instruments. He explained that the statistical methods currently used in QoL research (i.e., classical test theory and item response theory [IRT]) are globally satisfactory for developing accurate and reliable QoL questionnaires and that these methods will likely help resolve some problems encountered with the interpretation and/or analysis of QoL outcomes. However, he emphasized that these problems cannot be the subject of a purely methodological/statistical response and that the concept of QoL should also be analyzed. While it was recognized that QoL instruments have globally captured major trends in individuals' health and have been useful, the concept of QoL in the available instruments can be criticized because it is too closely associated with physical and cognitive functioning. This narrow view of QoL often excludes a key element: the feelings and emotions of individuals. The hardest challenge likely to be faced in the coming years is to work on these conceptual aspects rather than on methodological/statistical problems. This important message has guided all the previous conferences, in which discussions about statistical methods have always prioritized the conceptual underpinnings of the topic under investigation.

Three main methodological/statistical topics were covered during the current conference: the treatment of missing data (MD), the development of item banking and computerized adaptive testing (CAT) and the detection/understanding of response shift (RS).

The treatment of missing data

M Mercier insisted at the beginning of this session that MD are a common problem in all

 10.1586/14737167.2014.873704 ISSN 1473-7167

types of medical research and that for QoL measures, these MD are common [1], especially in longitudinal studies of patients with cancer. JB Hardouin recalled that QoL questionnaires can be partially (only a few items) or completely missing (the whole questionnaire). He then described the three missing data mechanisms: missing completely at random: the MD is independent of the patient's QoL; missing at random: the MD may depend on covariates describing the patients' or the items' characteristics and missing not at random: the MD depends on the patient's (unobserved) QoL. JB Hardouin indicated that the most common imputation method used in medical literature (i.e., personal mean score or simple imputation method) should be avoided and replaced by relevant imputation methods that are currently available for handling missing data. One explanation for this misuse could be the ease of performing simple imputation in comparison with other methods that imply relatively complex programming software such as R, SAS and Stata. Then, he described an alternative appropriate method for handling missing data using the Rasch model, especially when the missing data are suspected to be missing not at random [2]. Finally, A Anota spoke about the recent interest in the survival analysis of time to QoL score deterioration in the field of oncology [3]. She recalled that the management of missing data is crucial in this situation and especially when the whole questionnaire is missing. She demonstrated that the results of these analyses are different according to the definition of time to QoL deterioration and the definition of the reference time, suggesting that the next step should be to reach a consensual definition of the event of QoL score deterioration similar to RECIST criteria for tumor measurement.

In response to these presentations, M Mercier told the audience that these statistical analyses can only be seen as 'palliative strategies', and they never replaced the 'preventive strategies' related to data collection. According to her, these preventive strategies should be considered the first priority when QoL data are collected (e.g., appropriate time to collect QoL data, auxiliary information to explain mechanisms of missing QoL data, etc.).

Item banking & computerized adaptive testing

F Guillemin explained in his presentation that QoL questionnaires are not always adapted to all clinical and socio-cultural situations or patient characteristics. Providing shorter or adaptive QoL questionnaires appeared to be appropriate in clinical practice. The first speaker, M Mesbah, gave a review of IRT assumptions (local independence, unidimensionality, local sufficiency, monotonicity, invariance and absence of differential item functioning [DIF]) used in the selection of items. The second speaker, P Michel, described the main principles of developing CAT, which consists of administering questions dynamically tailored to each individual from an item bank. An item bank is a collection of items that measure the same concept. By fitting an IRT model for polytomous data to the item responses, all the items of the bank are calibrated and then used in a CAT. The obvious advantages of CAT were recalled, such as a minimized number of questions and items tailored to the characteristics of an individual. However, several constraints of CAT were also discussed, among which the two most important were: CAT is a very expensive approach because item calibration requires a large number of patients, and the involvement of important human and computer resources that are rarely available in health centers. These research perspectives are important, especially when taking into account the multidimensionality of QoL. Interestingly, Bacci and Bartolucci recently proposed a multidimensional latent classes Rasch model, which may take into account the correlations between the different dimensions of QoL [4].

Presentations by J Coste and A Rouquette covered DIF in QoL data, that is, the probability of item response differences across comparison groups after conditioning on the QoL level. Although DIF is commonly studied in QoL questionnaires to understand whether between-group QoL score differences are true QoL differences or are due to DIF, this notion is rarely investigated in the context of traditional clinical and biological outcomes. The speakers questioned the significance of DIF detection in QoL data: when should the QoL score difference be interpreted as a measurement bias or as the true difference between the groups? The audience wondered about the consequences of DIF detection and whether DIF should be integrated into the interpretation of QoL data. P Auquier informed the audience that the identification of a 'too high-level' requirement for QoL measures can also be used by the medical community and by decision makers as a justification for not using these measures, reinforcing what they denounce: a lack of impact on clinical management and decision-making [5].

Response shift

F Bonnetain outlined that QoL changes found in longitudinal studies are difficult to interpret. Are these changes due to a true change of the QoL level or to respondents' changing standards, values or conceptualizations [6]? This latter phenomenon, also called RS, is usually considered to be a measurement bias that changes with the time of measurement in longitudinal research. Classically, three types of RS have been distinguished: changes in internal standards of measurement (recalibration), changes in the priority (i.e., importance) of the component domains of the target construct (reprioritization) and redefinition of the target construct (reconceptualization) [7]. Presentations by MGE Verdam, M Boucekine and A Guilleux covered three technical aspects of RS detection: structural equation modeling (SEM), random forest models and IRT models, respectively [7–9]. MGE Verdam explained that SEM offers a useful statistical approach for RS detection because it allows for the operationalization of the three different types of RS. Moreover, compared with the then-test approach (i.e., the most common approach to the detection of RS), the advantages of SEM are that it does not require an extra assessment and that it can be more easily applied to multiple time measurements. MGE Verdam also showed that

SEM can be used to test the underlying assumptions of the then-test (i.e., the so-called 'Recall assumption', 'Consistency assumption' and 'Recalibration assumption') [10]. M Boucekine illustrated how the variable importance score provided by the random forest method can be used to identify the reprioritization type of RS [8]. Finally, A Guilleux compared the SEM approach proposed by Oort [9] and the IRT models using the longitudinal generalized partial credit model and the longitudinal partial credit model to detect the reprioritization and recalibration components of RS. She reported discordant results between the two methods: different items are affected by recalibration with SEM and IRT, and more items are affected by reprioritization with IRT. The advantages and limitations of each method have been clearly discussed. The participants considered that it would be premature to conclude which method is best for detecting an RS. The variety of methods developed illustrates the complexity and difficulty of detecting and measuring an RS. How to integrate the RS in the interpretation of QoL scores remains a challenge for future research. The speakers questioned the clinical relevance of RS detection, particularly if RS should be interpreted as a measurement bias. F Bonnetain told the audience that clinicians and particularly oncologists are always perplexed by an observed difference in QoL scores between two groups of treated patients. These clinicians need help to interpret the meaning of these differences and distinguish the part of the true change and the part of change related to RS. P Auquier concluded the discussion with a comment on RS: the true change of the QoL level can also be considered to be directly linked to the respondents' changing standards or values. In this sense, RS cannot be considered in terms of measurement bias [11]. Finally, counter-intuitive findings (e.g., the same QoL level before and after the occurrence of disease) can be paradoxical only for experts and not for patients, confirming the discrepancy between the views of patients and professionals.

Authors contribution

Drafting and writing of the Meeting Report was completed by all of the authors.

Acknowledgements

This work was supported by the French National Platform of Quality of Life and Cancer, the French Institut de Recherche en Santé Publique (IRESP CUD-QV, Concepts, Usages et Déterminants en Qualité de Vie) and the French Statistical Society. The authors would like to thank all the participants of this conference.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending or royalties.

No writing assistance was utilized in the production of this manuscript.

References

1. Quinquis L, Audureau E, Pouchot J, et al. Non-response, incomplete and inconsistent responses to self-administered health-related quality of life measures in the general population: patterns, determinants and impact on the validity of estimates – a population-based study in France using the MOS SF-36. Health Qual Life Outcomes 2013;11:44
2. De Bock E, Hardouin JB, Blanchin M, et al. Rasch-family models are more valuable than score-based approaches for analysing longitudinal PRO with intermittent missing data. Stat Meth Med Res 2013;In press
3. Bonnetain F, Dahan L, Maillard E, et al. Time until definitive quality of life score deterioration as a means of longitudinal analysis for treatment trials in patients with metastatic pancreatic adenocarcinoma. Eur J Cancer 2010;46:2753-62
4. Bacci S, Bartolucci F. A multidimensional latent class Rasch model for the assessment of the health-related quality of life. Christensen KB, Mesbah M, Kreiner S, editors. Rasch models for health sciences. John Wiley & Sons, Inc; Hoboken, NJ, USA: 2012. p. 199-222
5. Boyer L, Auquier P. The lack of impact of quality-of-life measures in schizophrenia: a shared responsibility? Pharmacoeconomics 2012;30:531-2.author reply 532-3
6. Dabakuyo TS, Guillemin F, Conroy T, et al. Response shift effects on measuring post-operative quality of life among breast cancer patients: a multicenter cohort study. Qual Life Res 2013;22:1-11
7. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. Soc Sci Med 1999;48:1507-15
8. Boucekine M, Loundou A, Baumstarck K, et al. Using the random forest method to detect a response shift in the quality of life of multiple sclerosis patients: a cohort study. BMC Med Res Methodol 2013;13:20
9. Oort FJ. Using structural equation modeling to detect response shifts and true change. Qual Life Res 2005;14:587-98
10. Verdam MGE, Oort FJ, Visser MRM, Sprangers MAG. Response shift detection through then-test and structural equation modeling: decomposing observed change and testing tacit assumptions. Neth J Psychol 2012;67:58-67
11. Oort FJ, Visser MR, Sprangers MA. Formal definitions of measurement bias and explanation bias clarify measurement and conceptual perspectives on response shift. J Clin Epidemiol 2009;62:1126-37

Affiliations

Laurent Boyer
Aix-Marseille Université, EA 3279 Research Unit, Marseille, France

Karine Baumstarck
Aix-Marseille Université, EA 3279 Research Unit, Marseille, France

Pierre Michel
Aix-Marseille Université, EA 3279 Research Unit, Marseille, France
and
FRE3529 The statistical team of the Mathematical Institute of Luminy, Marseille, France

Mohamed Boucekine
Aix-Marseille Université, EA 3279 Research Unit, Marseille, France
and
FRE3529 The statistical team of the Mathematical Institute of Luminy, Marseille, France

Amelie Anota
The French National Platform of Quality of Life and Cancer, France
and
Methodology and quality of life unit in oncology, EA 3181, CHRU Besançon, Besançon, France

Franck Bonnetain
The French National Platform of Quality of Life and Cancer, France
and
Methodology and quality of life unit in oncology, EA 3181, CHRU Besançon, Besançon, France

Joel Coste
Biostatistics and Epidemiology Unit, Hôtel Dieu Hospital, AP-HP, Paris, France

Bruno Falissard
Inserm U669, Mental Health and Public Health, Paris, France

Alice Guilleux
EA4275-SPHERE "bioStatistics, Pharmacoepidemiology and Human sciEnces Research tEam", Nantes, France

Jean-Benoit Hardouin
EA4275-SPHERE "bioStatistics, Pharmacoepidemiology and Human sciEnces Research tEam", Nantes, France

Anderson Loundou
Aix-Marseille Université, EA 3279 Research Unit, Marseille, France

Mariette Mercier
The French National Platform of Quality of Life and Cancer, France
and
EA3181, University of Franche-Comte, Besançon, France

Mounir Mesbah
Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie Paris VI, 175 rue du Chevaleret, 75013 Paris, France

Alexandra Rouquette
Biostatistics and Epidemiology Unit, Hôtel Dieu Hospital, AP-HP, Paris, France
and
Inserm U669, Mental Health and Public Health, Paris, France

Veronique Sebille
EA4275-SPHERE "bioStatistics, Pharmacoepidemiology and Human sciEnces Research tEam", Nantes, France

Mathilde GE Verdam
Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, The Netherlands
and
Department of Medical Psychology, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands

Badih Ghattas
FRE3529 The statistical team of the Mathematical Institute of Luminy, Marseille, France

Francis Guillemin
The French National Platform of Quality of Life and Cancer, France
and
EA4360 APEMAC Lorraine University, Paris Descartes, Nancy, France

Pascal Auquier
Aix-Marseille Université, EA 3279 Research Unit, Marseille, France
and
The French National Platform of Quality of Life and Cancer, France