



The use of electronic PROMs provides same outcomes as paper version in a spine surgery registry. Results from a prospective cohort study

Francesco Langella¹ · Paolo Barletta¹ · Alice Baroncini¹ · Matteo Agarossi¹ · Laura Scaramuzzo¹ · Andrea Luca¹ · Roberto Bassani¹ · Giuseppe M. Peretti^{1,2} · Claudio Lamartina¹ · Jorge H. Villafañe³ · Pedro Berjano¹ 

Received: 2 April 2020 / Revised: 23 March 2021 / Accepted: 3 April 2021
© The Author(s) 2021

Abstract

Background and Purpose Patient-Reported Measured Outcomes (PROMs) are essential to gain a full understanding of a patient's condition, and in spine surgery, these questionnaires are of help when tailoring a surgical strategy. Electronic registries allow for a systematic collection and storage of PROMs, making them readily available for clinical and research purposes. This study aimed to investigate the reliability between the electronic and paper form of ODI (Oswestry Disability Index), SF-36 (Short Form Health Survey 36) and COMI-back (Core Outcome Measures Index for the back) questionnaires.

Methods A prospective analysis was performed of ODI, SF-36 and COMI-back questionnaires collected in paper and electronic format in two patients' groups: Pre-Operatively (PO) or at follow-up (FU). All patients, in both groups, completed the three questionnaires in paper and electronic form. The correlation between both methods was assessed with the Intraclass Correlation Coefficients (ICC).

Results The data from 100 non-consecutive, volunteer patients with a mean age of 55.6 ± 15.0 years were analysed. For all of the three PROMs, the reliability between paper and electronic questionnaires results was excellent (ICC: ODI = 0.96; COMI = 0.98; SF36-MCS = 0.98; SF36-PCS = 0.98. For all $p < 0.001$).

Conclusions This study proved an excellent reliability between the electronic and paper versions of ODI, SF-36 and COMI-back questionnaires collected using a spine registry. This validation paves the way for stronger widespread use of electronic PROMs. They offer numerous advantages in terms of accessibility, storage, and data analysis compared to paper questionnaires.

Keywords Clinical outcome · Spine registry · Validation · Electronic data collection · Quality of life score

✉ Pedro Berjano
pberjano@gmail.com

Francesco Langella
francesco.langella.md@gmail.com

Paolo Barletta
paolo.barletta@grupposandonato.it

Alice Baroncini
alice.baroncini@gmail.com

Matteo Agarossi
matteo.agarossi@gmail.com

Laura Scaramuzzo
scaramuzzolaura@gmail.com

Andrea Luca
aluca.md@gmail.com

Roberto Bassani
r.bassani.spine@gmail.com

Giuseppe M. Peretti
giuseppe.peretti@unimi.it

Claudio Lamartina
c.lamartina@chirurgiavertebrale.net

Jorge H. Villafañe
mail@villafane.it

- 1 IRCCS Istituto Ortopedico Galeazzi, Milano, Italy
- 2 Department of Biomedical Sciences for Health, University of Milan, Milan, Italy
- 3 IRCCS Fondazione Don Carlo Gnocchi, Milano, Italy

Introduction

Spinal disorders represent a common clinical condition and leads to a substantial social and economic burden: yearly, up to 15% of the population will suffer a first-time back pain episode and up to 80% of patients will experience a recurrence episode [1]. Even if the complication rate is decreasing over time, the late-onset complications after complex spine surgery are still considerable [2]. Beyond the social and economic influence of these factors [3], they can lead to a decrease in the patients' quality of life. The evaluation of this last aspect should always be taken into account in the pre- and post-operative assessment of the patients.

Patient-Reported Outcome Measures (PROMs) offer a useful tool to assess subjective clinical data such as pain or quality of life [4]. Still, it is fundamental to combine these parameters with objective data such as surgical data and clinical outcomes [5]. Registries that systematically collect data regarding clinical practice, users' safety data, and PROMs are thus an exceptional opportunity to monitor the impact and value of surgical procedures and to conduct research on the factors influencing the results and complication rate of different surgical techniques [6, 7].

The development of a clinical registry allows merging a large amount of data [8], and traditional paper-based systems seem expensive and inefficient compared to electronic ones [9]. The electronic administration of PROMs allows an easier collection and management of a considerable volume of data, thus facilitating the workflow. Furthermore, an electronic format also permits the collection of all data in a single database and makes them readily available for extraction and further analysis. To achieve the goal of merging all the data regarding a patient in one electronic dataset, a validation of the electronic version of PROMs is required, as factors like the graphic organization of the elements on a screen or the impossibility to skip questions may alter the way patients answer the questionnaire [9].

The study aimed to assess the reliability and the agreement between PROMs administered via an electronic-based method and the data collected using a paper-based format.

Material and methods

This project was based on the retrospective analysis of patients prospectively enrolled in a spinal surgery registry: SpineReg [10]. The accuracy, reliability, and validity of the electronic-based data were evaluated through the comparison with the paper-based data collection, which is considered the gold standard.

The study was conducted at a single-centre and included the analysis of patients' reported outcomes in subjects prior to or after spine surgery. The inclusion criteria were age ≥ 18 years, both genders, and the capability to read and understand the Italian language. Patients who were unable to understand and answer questionnaires independently were excluded from the study. The study protocol was in accordance with the Helsinki Declaration of 1957 as revised in 2000. The procedures followed the ethical standards of the responsible committee on human experimentation and was approved by the ethics committee of our Institution (second amendment to the SPINEREG protocol issued on 13/04/2016). The project was supported with funds from the Italian Ministry of Health (project code CO-2016-02,364,645). All patients gave their written informed consent for the participation in this study.

The sample size (100 subjects) was determined based on the study conducted by G. Rankin et al. [11] who established the minimum number of cases required to test a musculoskeletal disorders questionnaire. All patients awaiting surgery or in follow-up at our institution and who voluntarily agreed to participate in this study were stratified in two groups: the Follow-Up group (FU), who received the questionnaires during an outpatient appointment for the routine follow-up, and the Pre-Operative group (PO), which filled out the questionnaires at the hospital admission, before surgery. Patients in the FU group filled in the questionnaires in paper form before (T1), and electronic form after the clinical examination (T2). The electronic version was administered on a tablet with the help of a supervisor, whose only role was to provide instructions on how to operate the tablet for patients who were not familiar with the device. Further instruction on how to interpret the question (e.g. Numeric Rating Scale—NRS: 0 = no pain, 10 = worst pain imaginable) was embedded in the questionnaire. Unlike the paper form, the electronic version only showed one item per screen, but the users could scroll forwards or backwards. Questionnaire explanations were available on the top of the screen during the entire completion of the questionnaire. The current study was performed on a tablet. However, the software maintains a similar layout on multiple devices such as PC, tablets and smartphones; except for the COMI pain scales, for which it was neither possible to add anchors at each end of the scale nor (in the smartphone presentation) to present the scale horizontally, due to the small screen size. An example of how questionnaires appear to the patients is shown in Fig. 1. Patients in the PO group completed the electronic questionnaires first, using a tablet and under supervision (T1). After one hour, the patients completed the same PROMs in paper format (T2). The questionnaires in electronic and paper versions were filled out on

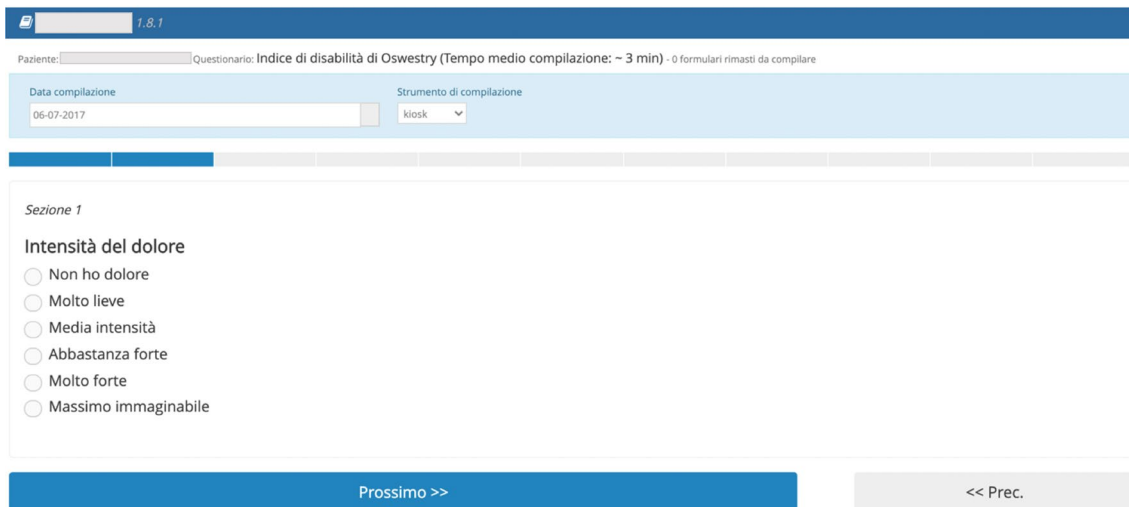


Fig. 1 User interface: item presentation. The picture shows the item presentation extracted from the ODI questionnaire (Italian version) of the electronic format. The patients can go forward (Prossimo>>) or backward (Prec.<<). A loading line (blue segmented bar) displays the percentage of questionnaire completion. The subjects can visualize the mean time (Tempo medio di compilazione ~3 min) and the

number of the questionnaires necessary to complete the survey (0 formulari rimasti da compilare). Paziente, Patient's name and surname; Questionario, Questionnaire; Data di compilazione, Compilation date; Strumento di compilazione, Instrument used to fill the questionnaire (tablet, e-mail, outpatients' interview, kiosk, phone or paper format)

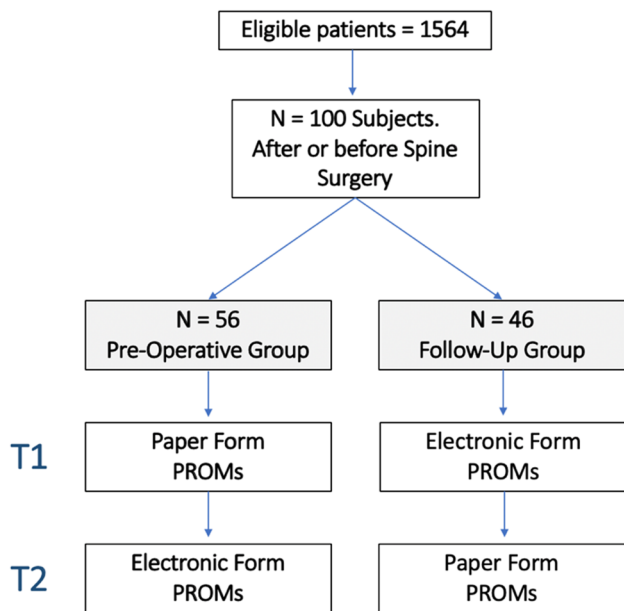


Fig. 2 Schematic representation of the study. The first evaluation by either paper or a tablet (electronic format) took place in the day of admission in hospital (Pre-Operative Group) and during follow-up visit (Follow-Up Group). The second evaluation either by a tablet or paper after 1 h. PROMs=Patient-Reported Outcome Measures [Short Form-36 (SF-36), Oswestry Disability Index (ODI) and Core Outcome Measures Index for the back (COMI-back)]

the same day in order to eliminate any score variations due to a change in the patient's clinical condition. A schematic representation of the study is presented in Fig. 2.

The recruited subjects were asked to fill in three questionnaires using the already validated Italian versions of the Oswestry Disability Index (ODI) [12], the Short Form Health Survey (SF-36) [13] with Physical Component Score (PCS) and Mental Component Score (MCS) and the Core Outcome Measures Index for the back (COMI-back) [14], [15]. These three questionnaires are routinely used to assess the disability associated with pain and examine the general health status of the patients [16–18].

Numerical data were expressed as Mean ± Standard Deviation (SD). The items correlation between electronic and paper version was estimated using the Gamma Correlation Coefficient. The correlation was defined excellent for $\gamma = 1.0$, strong for $0.3 \leq \gamma < 1$, moderate for $0.09 \leq \gamma < 0.3$ and weak for $0.01 \leq \gamma < 0.09$. To assess the reliability between Likert type item's response in paper and electronic format, the linear weighted Kappa statistics was used. The levels of Kappa statistics were defined as follows: $0.00 < K < 0.20$ poor or slight agreement; $0.21 < K < 0.40$ fair; $0.41 < K < 0.60$ moderate; $0.61 < K < 0.80$ substantial or good; $0.81 < K < 1.00$ very good or perfect [19].

The paired t-test for related samples was used to compare normally distributed parameters. The Mann–Whitney t-test was performed to determine the differences in continuous variables between the two cohorts, while the Chi-square test (χ^2) was used for the categorical variables. Non-normal variables were compared with the two-tailed Wilcoxon Range Test for dependent variables.

In accordance with previous studies on the same topic, an ICC value greater or equal to 0.90 signified an excellent

reliability between the electronic and paper forms of the PROMs. [20, 21]

To determine the amount of variation in the measurement errors for the electronic and paper format, the Standard Error of Measurement [$SEM_{\text{agreement}} = \sqrt{\sigma_r^2 + \sigma_{pt}^2}$, where σ_{pt}^2 represents the variance due to systematic differences between the two types of questionnaire administration (electronic and paper); and σ_r^2 represents the residual variance (namely the part of the variance which cannot be attributed to specific causes)] was calculated, which also accounted for interrater variation in order to provide an agreement measure. Furthermore, the minimal detectable change (MDC_{95}) [$MDC = 1.96 \times \sqrt{2} \times SEM$] was calculated to estimate the size of any change that was not likely due to measurement error [22]. The Bland–Altman analysis graphically showed the agreement by plotting the difference between paper and electronic scores against their mean. Limit of Agreement (LOA) was also used to estimate the absolute agreement between test–retest of the different questionnaires [23]. Furthermore, the whole sets of agreement and correlation tests between paper and electronic format were calculated for PO and FU groups. The statistical analysis was performed with SPSS (SPSS Inc. Version 22.0, Armonk, NY: IBM Corp.) and conducted with a 95% confidence interval (CI). The degree of statistical significance was set at p value < 0.05 .

Results

From January 2017 to October 2019, 1564 patients requiring spine surgery were recorded in the electronic registry and were eligible for the study. Of these, 100 volunteered to participate in the study, gave their informed consent and were enrolled in the project (mean age 55.6 ± 14.9 years). Among these patients, 59% were women and 41% were men. Forty-six patients were enrolled in the PO group, and 56 in the FU group. The follow-up visits were performed 3 to 24 months after surgery. No statistically significant differences, in terms of age and gender, were observed between the eligible population and the patients enrolled in the study.

There were no significant differences in the demographics among the PO and FU group (female PO 59.3%, FU = 58.7%, $p = 0.954$; age PO 54.0 ± 15.9 , FU 57.5 ± 13.8 , $p = 0.294$). The parameters of evaluation of the quality of life and disability in the PO group were as follows: ODI 41.3 ± 17.3 , SF-36 MSC 46.7 ± 12.1 , SF-36 PCS 35.4 ± 8.2 , COMI-back 7.1 ± 1.9 . The results of the same questionnaires in the FU group were ODI 30.8 ± 18.6 , SF-36 MSC 48.8 ± 9.5 , SF-36 PCS 37.5 ± 9.2 , COMI-back 4.2 ± 2.2 .

The values of the Gamma coefficients showed a correlation between paper and electronic form ranging from good to excellent (Gamma minimum value = 0.91, maximum

value = 1.00). The weighted Kappa coefficients showed a significant positive agreement between paper and electronic form for all examined items (Kappa minimum value = 0.67, maximum value = 0.95).

The reliability between the results of paper and electronic questionnaires was excellent. The ICC values ranging from 0.963 (ODI, overall sample) to 0.982 (COMI-back, overall sample). The test–retest reliability was high for the PO and FU groups ($p < 0.001$ for both). The PO and FU groups mean absolute difference, SEM and MDC_{95} for paper and electronic reliability are reported in Table 1. The 95% limit of agreement ranged from -9.72 to 9.04 for the ODI, from -0.82 to 0.75 for the COMI-back, from -4.01 to 4.45 for SF-36 MCS, and from -3.38 to 3.37 for SF-36 PCS. The mean \pm SD and LOA of Pre-Operative and Follow-up groups are summarized in Table 1.

The Bland–Altman graph (Fig. 3) showed that the mean differences were -0.337 for the ODI, -0.036 for the COMI-back, 0.244 for the SF-36 MCS and -0.012 for the SF-36 PCS. The differences between the two acquisition formats were plotted around the zero line for all the questionnaires. No systematic bias comparing electronic and paper form was detected.

Discussion

Several correlation studies proved the correlation between paper and electronic format of PROMs [9]. Our results were consistent with scientific literature examining the reliability and consistency of the electronic versions of the PROMs, along with result comparability with the paper form [24]. The Gamma coefficients denoted a high correlation between the two methodologies used for filling out the questionnaires. Furthermore, the weighted Kappa Coefficient showed a good to perfect reliability between paper and electronic version.

The comparison between the results of paper and electronic questionnaires revealed an excellent test–retest reliability for all questionnaires. The SEM for the overall scores was acceptable for all questionnaires: 3.39/100 points for ODI, 0.28/10 for COMI-back, 1.53/100 for SF-36 MCS and 1.21/100 for SF-36 PCS. Furthermore, the Bland–Altman graph showed limits of agreement values comparable with MDC_{95} of our study; both were smaller if compared with MCIC values of the literature for all the questionnaires analysed (ODI = 15U [12], COMI-back = 3U [16], SF-36 = 4.9U [25]). Thus, the electronic form can be used for the same clinical purpose as the paper form. The mean bias for paper and electronic forms was always plotted around zero, confirming the absence of a systematic error. Since the two analysed groups were uniform and there were no systematic differences between electronic and paper collection methods, these can be used interchangeably.

Table 1 Reliability and Agreement analysis

Questionnaire	ICC	95% Confidence interval		Limit of Agreement Questionnaires Scores				SEM	MDC _{95%}	
		Lower bound	Upper bound	Paper version	Electronic version	Bias	95% LOA			
							Lower			Upper
ODI PO	.998	.996	.999	38.4±15.9	38.8±16.3	-0.37±1.0	-2.33	1.59	2.059	5.706
ODI FU	.928	.873	.959	30.1±18.5	30.4±18.1	-0.30±7.0	-14.02	13.42	5.181	14.360
ODI (Overall)	.963	.946	.975	34.6±17.6	34.±17.6	-0.34±4.7	-9.72	9.04	3,385	9,383
COMI-back PO	.999	.998	.999	5.7±1.6	5.7±1.8	-0.02±0.1	-0.22	0.18	0.131	0.364
COMI-back FU	.955	.920	.975	3.6±1.9	3.7±1.9	-0.05±0.6	-1.23	1.13	0.874	2.443
COMI-back (overall)	.982	.973	.988	4.7±2.1	4.7±2.0	-0.04±0.4	-0.82	0.75	0,285	0,789
SF-36 MCS PO	.993	.988	.996	46.7±11.9	46.7±12.1	-0.02±1.4	-2.76	2.72	1.021	2.830
SF-36 MCS FU	.957	.924	.976	49.3±9.6	48.8±9.5	0.51±2.8	-4.98	5.99	3.149	8.729
SF-36 MCS (overall)	.981	.971	.987	47.9±10.9	47.6±10.9	0.22±2.1	-4.01	4.45	1,526	4,229
SF-36 PCS PO	.993	.989	.996	35.5±8.5	35.4±8.2	0.13±0.9	-1.63	1.89	0.941	2.608
SF-36 PCS FU	.968	.942	.982	37.4±8.9	37.5±9.2	-0.15±2.3	-4.66	4.36	1.795	4.976
SF-36 PCS (overall)	.981	.971	.987	36.3±8.6	36.3±8.6	0.0006±1.7	-3.38	3.37	1,212	3,359

ODI, Oswestry Disability Index; COMI-back, Core Outcome Measures Index for the back; SF-36 MCS, Short Form-36 Mental Component Score; SF-36 PCS, Short Form-36 Physical Component Score; ICC, Intraclass Correlation Coefficient ICC Using Single-Rating, Absolute-Agreement, 2-Way Mixed-Effects Model, single measures; IC 95%, Confidence Interval at 95%; Bias, difference between paper and electronic version; 95% LOA, 95% limits of agreement; SEM, Standard Error of Measurement; MDC_{95%}, Minimal Detectable Change at 95% Confidence Interval. Data are expressed as Means ± Standard Deviation

Having assessed the absence of significant differences between the two methods, the authors believe that the electronic will over time substitute the paper form. It is possible to make this prediction since the distribution, collection and compilation of PROMs within electronic registers entails obvious advantages regarding logistics and data usability [26]. Several advantages can be obtained with the systematic use of electronic formats. Patient scores can be calculated in real-time and be easily monitored and shared with the other members of the research team. An electronic-based data collection allows patients to complete the questionnaires remotely, thus reducing the time spent in hospital facilities. This represents a considerable advantage to preserve the social connections while promoting physical distancing [27]. The distribution of electronic questionnaires is also favoured by the constant increase in users of electronic devices (smartphones, tablets or personal computers). According to epidemiologic surveys, 80% of people in Europe have internet access, with peaks of 88% and 87%, respectively, in the UK and Spain [28]. In Italy, 76.1% of the population has an internet connection; this percentage grows to over 90% among young people (15–24 years), but the number of internet users among 65–74 year-olds is also increasing in the last years (up to 41.9%) (ISTAT 2019—Italian Statistics Institute). Despite these encouraging numbers, clinicians cannot yet assume that all patients have an internet connection and are familiar enough with electronic devices to fill the questionnaires out remotely. It is essential to discuss these topics during office visits and

provide patients with the support they may require (e.g. written instructions on how to fill out the questionnaires online, provide a user-friendly graphic interface, provide a contact in case of further help) and offer alternative solutions such as phone calling interview. For this study, the patients were aided by supervisors to optimize the tablet use.

The advantages of PROMs collection in electronic format can lead to a further spreading in the use of this tool. It is established that patient-reported outcomes have a considerable clinical value and, combined with the objective clinical and surgical data, allow the clinician to obtain a complete and accurate understanding of a patient's condition [29]. Stored on a registry, all these data are readily available for use in personalized medicine [30]. This is particularly significant for spine conditions, where numerous available surgical techniques and the differences in surgical strategies open a wide range of different scenarios [31, 32]. The storage and analysis of the outcomes of previously treated patients may help defining a surgical approach tailored on the needs of a specific patients' group [33–35]. In 2019, Ghogawala et al. [36] examined how clinical registries might be used to generate new evidence to support a specific treatment option when comparable alternatives exist. In particular, the use of artificial intelligence can be aimed at building mathematical algorithms to approximate conclusions from real medical and patient-reported data to develop data-driven personalized care [37]. In the era of “Omic” data, with a large mixture of information coming from several medical fields (genomics, proteomics, nutrigenomics and phenomics),

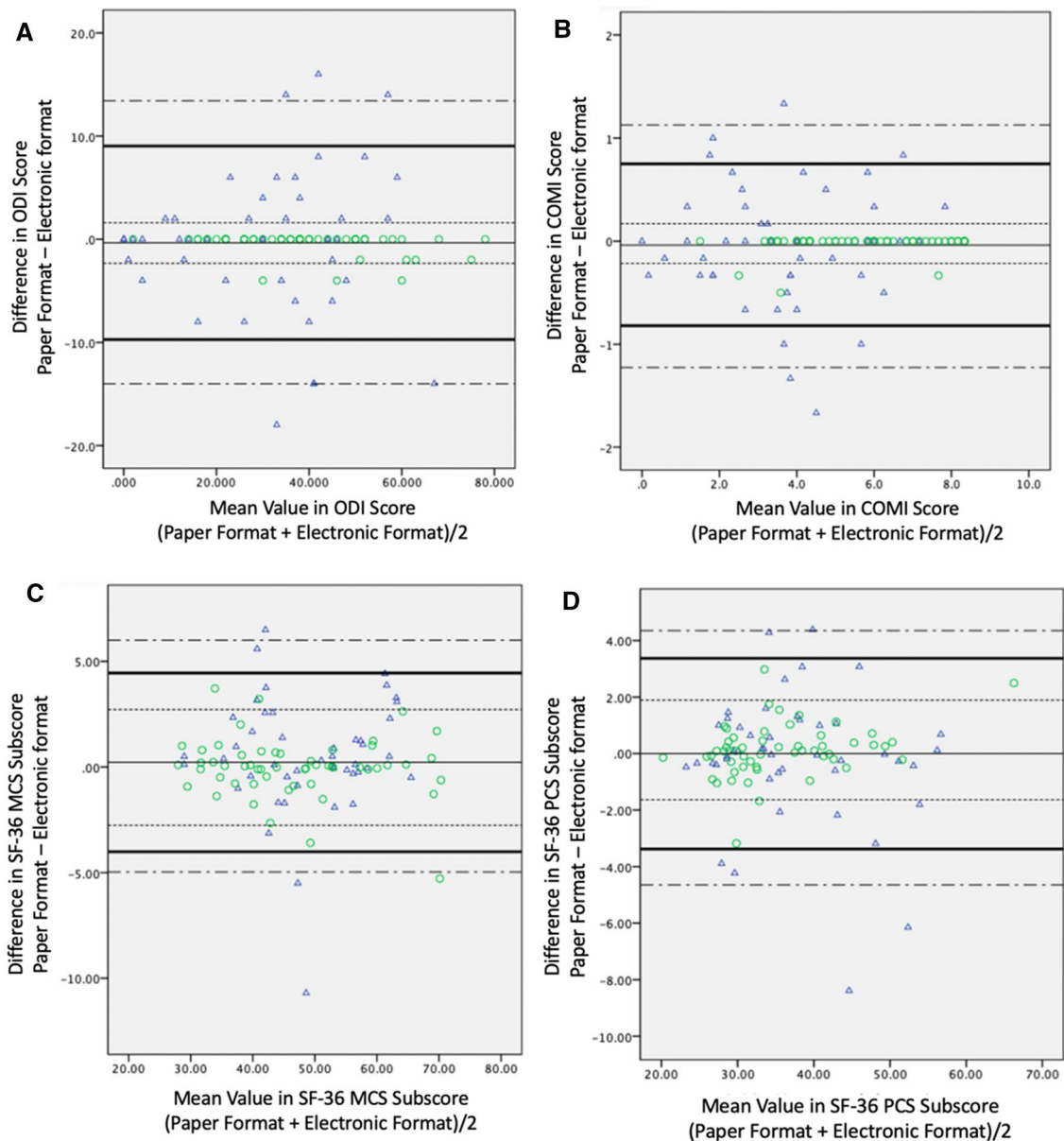


Fig. 3 Bland–Altman plot. Bland Altman plot of ODI (A), COMI-back (B), SF-36 MCS (C) and SF-36 PCS (D) assessed for electronic format and paper format. The bias (mean) between the two methods is marked by the full line (—), the overall upper and lower limits of agreement (LOA) by bold line (—), the Pre-Operative upper and lower LOA by the broken line (—) and the Follow-Up upper and

lower LOA by dash-dotted line (—●—). SF-36 MCS, Short Form-36 Mental Component Score; SF-36 PCS, Short Form-36 Physical Component Score; ODI, Oswestry Disability Index; COMI-back, Core Outcome Measures Index for the back. *Triangles*: Follow-up cases; *Circles*: Pre-Operative cases

the registries can make the difference in the capability to identify novel associations between biological and medical events [38]. An electronic registry allows the collection and storage of a large amount of records in a single database, permitting in turn easy access and extraction of the data for multiple purposes: follow-up and complication analysis, research purposes [39], quality control and risk modelling, establishing standards, share best practices, and build trust among institutions [40, 41].

PROMs have increasingly been raising interest in their predictive value. In a recent systematic review [42] predictive models using PROMs in spine surgery could be used for quantitative prediction of clinical outcomes. In particular, the authors were able to define a mathematical model capable of identifying several key predictors of surgical results such as age, sex, BMI, ASA score, smoking status, and previous spinal surgeries. Another recent study [25] showed that a specific pre-operative threshold for ODI and

other PROMs was capable of predicting a significant clinical improvement after surgery for adult spine deformity. These results highlight the usefulness of PROMs in pre-operative decision-making. Given the changing health-care environment, the electronic-based data collection systems seem essential and can boost the development of surgical registries.

Despite the encouraging conclusion of our study, several limitations should be taken into account evaluating the results. First, the lack of randomization and a low rate of eligible patients that accepted to participate may have introduced possible selection bias. The patients filled the questionnaires in paper and electronic form at two close moments in time: this entails a high reproducibility of mnemonic errors in completing the two questionnaires. A longer pause between the two formats, or repeated submissions of the questionnaire, may have solved this limitation. An ideal scenario for assessing the reliability of repeated measurements would have been to administer the questionnaires the second and third time in the same way and in the same time intervals (i.e. T1. Paper > > T2. Electronic > > T3. Electronic, or T1. Electronic > > T2. Paper > > T3. Paper). This methodological approach would potentially prove that the “intra-method” and “inter-method” variability does not increase when questionnaires are administered electronically.

A substantial limitation of the study is related to the methodological approach in the FU population. Specifically, the test and retest were performed before and after the clinical examination, potentially introducing a distortion effect through the clinical interview. Based on our data, it appears that the error of measurement was higher in the FU population. In particular, the LOAs for ODI were acceptable in the overall population, but excessively wide in the FU group. This means that the agreement of these measures was not sufficient to allow them to be interchangeable—an aspect that requires further investigation. A random allocation to set the order in which the PROMs were administered (paper first or digital first), and the lack of intervention between the two administrations may have addressed some of the limits of the study.

Even considering the proofreading and the double entry verification, the additional steps of the paper version related to the transcription process may result in a potential source of bias.

Furthermore, the influence of a supervisor while the patients filled the questionnaire may have introduced a bias. The presence of a supervisor and his or her attitude towards PROMs has been proved to have an influence on patients, most of all on the more fragile ones [43]. Further studies are required to investigate the effects of the presence of a supervisor on the patients’ perspective and to analyse how this reflects on studies results. An analysis of possible differences in the time required for the patients to fill out the different

versions of the questionnaires and on the preferred format was not performed. The fact that the PO and FU groups were administered the questionnaire in different succession may have also introduced a bias; however, a separate analysis between PROMs collected before surgery and at follow-up was not conducted. Considering the limitations of the study, our results only provide preliminary evidence. Further research is necessary to consolidate the acquired knowledge.

Conclusion

The result of our study showed excellent reliability and a significant agreement between the paper-based and the electronic-based data collection system of three relevant questionnaires in spine surgery, ODI, SF-36, and COMI-back. This validation supports the use of the electronic versions of these PROMs, which allows for quick accessibility of the data in the daily clinical practice and research purposes. Further studies will be necessary to consolidate the acquired knowledge and prove the efficacy of electronic registries to pave the way for data-driven personalized care.

Acknowledgements This study was funded by the Italian Ministry of Health (Project Code: CO-2016-02364645). The funders had no involvement in study design, collection, analysis and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication. The data that were used for this study were collected in the SpineReg electronic register (IOG SpineReg version 1.7.4, Deloitte, Milan, Italy).

Author’s contributions F. L. was involved in study design, data acquisition and analysis, draft and final approval of the article; Pa. Ba. was involved in study design, data acquisition and analysis, draft and final approval of the article; A. B. was involved in data interpretation, draft and final approval of the article; M. A. was involved in data acquisition, critical revision and final approval of the article; A. L. was involved in data acquisition, critical revision and final approval of the article; R. B. was involved in data acquisition, critical revision and final approval of the article; L. S. was involved in data acquisition, critical revision and final approval of the article; G. M. P. was involved in data acquisition, critical revision and final approval of the article; C. L. was involved in obtaining of funding, administrative support, data interpretation, critical revision and final approval of the article; J. H. V. was involved in data acquisition, critical revision and final approval of the article; Pe. Be. was involved in obtaining of funding, administrative support, study conception, data interpretation, critical revision and final approval of the article; Pa. Ba. and F. L. take responsibility for the work as a whole, from inception to finished article.

Declarations

Ethics approval and consent to participate Ethical approval for this study was asked for and waived by the local Medical Research Ethics Committee (Fourth Amendment to the SPINEREG Protocol, Issued on 10/10/2019). The study fell outside the remit of the law for Medical Research Involving Human Subjects Act and was approved by the local ethical committee. This study is registered in the US National

Library of Medicine (ClinicalTrials.gov Identifier: NCT03644407). All participants gave digital consent to participate in the present study.

Conflict of interest The authors declare that they have no competing interests.

Availability of data and material The datasets used and/or analysed in the present study are available from the corresponding author on reasonable request.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Hoy D, Brooks P, Blyth F, Buchbinder R (2010) The epidemiology of low back pain. *Best Pract Res Clin Rheumatol* 24:769–781. <https://doi.org/10.1016/j.berh.2010.10.002>
- Zanirato A, Damilano M, Formica M et al (2018) Complications in adult spine deformity surgery: a systematic review of the recent literature with reporting of aggregated incidences. *Eur Spine J* 27:2272–2284. <https://doi.org/10.1007/s00586-018-5535-y>
- Cortesi PA, Assietti R, Cuzzocrea F, et al (2017) Epidemiologic and economic burden attributable to first spinal fusion surgery: analysis from an Italian administrative database. *Spine (Phila Pa 1976)*. doi: <https://doi.org/10.1097/BRS.0000000000002118>
- Finkelstein JA, Schwartz CE (2019) Patient-reported outcomes in spine surgery: past, current, and future directions. *J Neurosurg Spine* 31:155–164. <https://doi.org/10.3171/2019.1.SPINE18770>
- Langella F, Villafañe JH, Damilano M, et al (2017) Predictive accuracy of surgimap surgical planning for sagittal imbalance: a cohort study. *Spine (Phila Pa 1976)* 42:E1297–E1304. doi: <https://doi.org/10.1097/BRS.0000000000002230>
- Berjano P, Langella F, Ismael M-F et al (2014) Successful correction of sagittal imbalance can be calculated on the basis of pelvic incidence and age. *Eur Spine J* 23(Suppl 6):587–596. <https://doi.org/10.1007/s00586-014-3556-8>
- Röder C, Müller U, Aebi M (2006) The rationale for a spine registry. *Eur Spine J* 15:S52–S56. <https://doi.org/10.1007/s00586-005-1050-z>
- Azad TD, Kalani M, Wolf T et al (2016) Building an electronic health record integrated quality of life outcomes registry for spine surgery. *J Neurosurg Spine* 24:176–185. <https://doi.org/10.3171/2015.3.SPINE141127>
- Park JY, Kim BS, Lee HJ et al (2019) Comparison between an electronic version of the foot and ankle outcome score and the standard paper version: a randomized multicenter study. *Medicine (Baltimore)* 98:e17440. <https://doi.org/10.1097/MD.00000000000017440>
- Ismael M, Villafañe JH, Cabitza F et al (2018) Spine surgery registries: hope for evidence-based spinal care? *J spine Surg (Hong Kong)* 4:456–458. <https://doi.org/10.21037/jss.2018.05.19>
- Rankin G, Stokes M (1998) Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil* 12:187–199. <https://doi.org/10.1191/026921598672178340>
- Monticone M, Baiardi P, Vanti C et al (2012) Responsiveness of the Oswestry Disability Index and the Roland-Morris Disability Questionnaire in Italian subjects with sub-acute and chronic low back pain. *Eur Spine J* 21:122–129. <https://doi.org/10.1007/s00586-011-1959-3>
- Apolone G, Mosconi P (1998) The Italian SF-36 health survey: translation, validation and norming. *J Clin Epidemiol* 51:1025–1036. [https://doi.org/10.1016/s0895-4356\(98\)00094-8](https://doi.org/10.1016/s0895-4356(98)00094-8)
- Mannion AF, Boneschi M, Teli M et al (2012) Reliability and validity of the cross-culturally adapted Italian version of the Core Outcome Measures Index. *Eur Spine J* 21. <https://doi.org/10.1007/s00586-011-1741-6>
- Neukamp M, Perler G, Pigott T et al (2013) Spine Tango annual report 2012. *Eur Spine J* 22(Suppl 5):767–786. <https://doi.org/10.1007/s00586-013-2943-x>
- Mannion AF, Porchet F, Kleinstück FS, et al (2009) The quality of spine surgery from the patient's perspective: part 2. Minimal clinically important difference for improvement and deterioration as measured with the Core Outcome Measures Index. *Eur Spine J* 18 Suppl 3:374–9. doi: <https://doi.org/10.1007/s00586-009-0931-y>
- Brazier JE, Harper R, Jones NMB et al (1992) Validating the SF-36 health survey questionnaire: New outcome measure for primary care. *Br Med J* 305:160–164. <https://doi.org/10.1136/bmj.305.6846.160>
- Fairbank JC, Pynsent PB (2000) The Oswestry Disability Index. *Spine (Phila Pa 1976)* 25:2940–52; discussion 2952. doi: <https://doi.org/10.1097/00007632-200011150-00017>
- Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
- Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 15:155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Weir JP (2005) Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J strength Cond Res* 19:231–240. <https://doi.org/10.1519/15184.1>
- Beckerman H, Roebroeck ME, Lankhorst GJ et al (2001) Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res*. <https://doi.org/10.1023/A:1013138911638>
- Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310
- Schröder ML, de Wispelaere MP, Staartjes VE (2019) Are patient-reported outcome measures biased by method of follow-up? Evaluating paper-based and digital follow-up after lumbar fusion surgery. *Spine J* 19:65–70. <https://doi.org/10.1016/j.spinee.2018.05.002>
- Copay AG, Glassman SD, Subach BR et al (2008) Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, medical outcomes study questionnaire short form 36, and pain scales. *Spine J* 8:968–974. <https://doi.org/10.1016/j.spinee.2007.11.006>
- Mayer CS, Williams N, Fung KW, Huser V (2019) Evaluation of research accessibility and data elements of HIV registries. *Curr HIV Res* 17:258–265. <https://doi.org/10.2174/1570162X17666190924195439>
- Bergman D, Bethell C, Gombojav N et al (2020) Physical distancing with social connectedness. *Ann Fam Med* 18:272–277. <https://doi.org/10.1370/afm.2538>
- Merkel S, Hess M (2020) The use of internet-based health and care services by elderly people in Europe and the importance

- of the country context: multilevel study. *JMIR aging* 3:e15491. <https://doi.org/10.2196/15491>
29. Basch E, Bennett A, Pietanza MC (2010) Use of patient-reported outcomes to improve the predictive accuracy of clinician-reported adverse events. *J Natl Cancer Inst* 103:1808–1810. <https://doi.org/10.1093/jnci/djr493>
 30. Stey AM, Russell MM, Ko CY et al (2015) Clinical registries and quality measurement in surgery: a systematic review. *Surg (United States)* 157:381–395. <https://doi.org/10.1016/j.surg.2014.08.097>
 31. Irwin ZN, Hilibrand A, Gustavel M, et al (2005) Variation in surgical decision making for degenerative spinal disorders. Part I: lumbar spine. *Spine (Phila Pa 1976)* 30:2208–13. doi: <https://doi.org/10.1097/01.brs.0000181057.60012.08>
 32. Lubelski D, Williams SK, O'Rourke C, et al (2016) Differences in the surgical treatment of lower back pain among spine surgeons in the united states. *Spine (Phila Pa 1976)* 41:978–86 . doi: <https://doi.org/10.1097/BRS.0000000000001396>
 33. Dui LG, Cabitza F, Berjano P (2018) Minimal important difference in outcome of disc degenerative disease treatment: the patients' perspective. *Stud Health Technol Inform* 247:321–325
 34. Haefeli M, Elfering A, Aebi M et al (2008) What comprises a good outcome in spinal surgery? A preliminary survey among spine surgeons of the SSE and European spine patients. *Eur Spine J* 17:104–116. <https://doi.org/10.1007/s00586-007-0541-5>
 35. Diebo BG, Henry J, Lafage V, Berjano P (2015) Sagittal deformities of the spine: factors influencing the outcomes and complications. *Eur Spine J* 24(Suppl 1):S3-15. <https://doi.org/10.1007/s00586-014-3653-8>
 36. Ghogawala Z, Dunbar MR, Essa I (2019) Lumbar spondylolisthesis: modern registries and the development of artificial intelligence. *J Neurosurg Spine* 30:729–735. <https://doi.org/10.3171/2019.2.SPINE18751>
 37. Campagner A, Berjano P, Lamartina C et al (2020) Assessment and prediction of spine surgery invasiveness with machine learning techniques. *Comput Biol Med* 121:103796. <https://doi.org/10.1016/j.combiomed.2020.103796>
 38. Alyass A, Turcotte M, Meyre D (2015) From big data analysis to personalized medicine for all: Challenges and opportunities. *BMC Med Genomics* 8:1–12. <https://doi.org/10.1186/s12920-015-0108-y>
 39. O'Connell S, Palmer R, Withers K et al (2018) Requirements for the collection of electronic PROMS either “in clinic” or “at home” as part of the PROMs, PREMs and effectiveness programme (PPEP) in Wales: a feasibility study using a generic PROM tool. *Pilot Feasibility Stud* 4:1–13. <https://doi.org/10.1186/s40814-018-0282-8>
 40. Nelson EC, Dixon-Woods M, Batalden PB et al (2016) Patient focused registries can improve health, care, and science. *BMJ* 354:1–6. <https://doi.org/10.1136/bmj.i3319>
 41. Larsson S, Lawyer P, Garellick G et al (2012) Use of 13 disease registries in 5 countries demonstrates the potential to use outcome data to improve health care's value. *Health Aff* 31:220–227. <https://doi.org/10.1377/hlthaff.2011.0762>
 42. Dietz N, Sharma M, Alhourani A et al (2018) Variability in the utility of predictive models in predicting patient-reported outcomes following spine surgery for degenerative conditions: a systematic review. *Neurosurg Focus* 45:E10. <https://doi.org/10.3171/2018.8.FOCUS18331>
 43. Fullerton M, Edbrooke-Childs J, Law D et al (2018) Using patient-reported outcome measures to improve service effectiveness for supervisors: a mixed-methods evaluation of supervisors' attitudes and self-efficacy after training to use outcome measures in child mental health. *Child Adolesc Ment Health* 23:34–40. <https://doi.org/10.1111/camh.12206>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.