Patient-perceived hand function measured can predict treatment for Dupuytren's disease

van den Berge, Bente A; Blom, Paula C A; Werker, Paul M N; Broekstra, Dieuwke C

Published in:
Plastic and Reconstructive Surgery

DOI:
10.1097/PRS.0000000000010515
10.1097/PRS.0000000000010515

Plastic and Reconstructive Surgery Advance Online Article

DOI: 10.1097/PRS.0000000000010515

Patient-perceived hand function measured can predict treatment for Dupuytren's disease

Bente A. **van den Berge**, MD[1]; Paula C.A. **Blom**, MD[1], Paul M.N. **Werker**, MD PhD[1]; Dieuwke C. **Broekstra**, PhD[1]

[1.] University of Groningen, University Medical Center Groningen, Dept. of Plastic Surgery, Groningen, the Netherlands

Corresponding author D.C. Broekstra University Medical Center Groningen Department of Plastic Surgery HPC BB81 P.O. box 30.001 9700 RB Groningen The Netherlands E-mail: d.c.broekstra@umcg.nl

Conflicting interests and financial disclosure PW was member of a Safety and Efficacy Review Board of Fidia ltd, Milan Italy. PW and DB are both member of the scientific advisory board of the Dutch Dupuytren Society and PW is member of the scientific advisory board of the International Dupuytren Society. This is not related to the submitted work. The other authors declare no potential conflicts of interest with respect to the content of this article. This research was funded by the University of Groningen and the C.&.W. de Boer foundation. The funding bodies had no influence on the design, conduct and analyses of this study.

Presentations: The results of this study have been presented at the International Dupuytren Conference 2021 (online), the EURAPS 2022 (Naples) and the IFSSH, IFSHT & FESSH Combined Congress 2022 (London).

Short running head:

Predicting surgery for DD with PROMs

ABSTRACT

Background: Web based patient-reported outcome measures (PROMs) could aid surgeons to remotely assess the need for examination and subsequent treatment of Dupuytren's disease (DD) patients. We studied whether the Unité Rhumatologique des affections de la Main (URAM) and the Michigan Hand Questionnaire (MHQ) could predict DD treatment.

Methods: In this prospective cohort study, we compared MHQ and URAM scores of treated patients with untreated patients. For the treatment group, we selected a score closest to one year before treatment. For controls we randomly selected a score. Additionally, we tested the predictive value of a one-year change score between 15 months and 6 weeks before treatment. The primary outcome measure was DD treatment.

The predictive value was determined using the Area Under the Curve (AUC). An AUC >0.70 was considered as good predictive ability, 0.70-0.50 as poor predictive ability and <0.50 as no predictive ability.

Results: We included 141 patients for the MHQ analysis and 145 patients for the URAM analysis. The AUC of the MHQ and URAM scores measured one year before treatment were 0.80 (95% CI 0.71-0.88) and 0.75 (95% CI 0.68-0.82), respectively. The one-year change score resulted in an AUC of <0.60 for both questionnaires.

Conclusions: Our results show that both the MHQ and URAM score measured around one year before treatment can predict treatment for DD. If future studies show that telemonitoring of DD patients with PROMs is also cost-effective, web-based PROMs could optimise patient care and treatment effectiveness of DD.

INTRODUCTION

Patients with Dupuytren's disease (DD) may suffer from an erratic disease course that is hard to predict.[1] Some patients are referred for treatment too early and some come so late that a good surgical result can no longer be obtained.[2–4] Ideally, patients should be examined regularly to assess the need for treatment.[3] Telemonitoring, or remote patient monitoring, could be a logistically and economically appealing method to keep DD patients on the radar, especially in times of reduced outpatient capacity, which was recently the case because of COVID-19.

Telemonitoring is emerging in many medical fields.[5] One way of telemonitoring is to ask patients to report (web based) disease-specific outcomes.[6] Such remotely assessed patient-reported outcome measure (PROM) scores are already implemented for monitoring chronic diseases such as chronic kidney disease, inflammatory bowel disease and rheumatoid arthritis.[7–9] In DD, clinicians and researchers utilise PROMs increasingly to evaluate the effect of treatment on patient-reported hand function. The Michigan Hand Questionnaire (MHQ)[10–12] and the Unité Rhumatologique des Affections de la Main (URAM) Scale[13] are among the most used PROMs to assess disability caused by DD.[14] They have shown validity and reliability in DD patients but are currently only applied to evaluate treatment results.[15] The use of these PROMs to remotely monitor the disease course of DD patients has not been explored.

In theory, the self-reported hand function of DD patients could be predictive of DD treatment. Hence, we hypothesised that monitoring patients with PROMs can be applied to predict the need of treatment for DD in the following year. In this study, we aimed to investigate the predictive ability of the MHQ and the URAM scale for treatment of DD.

METHODS

Study design and population

We used data of 261 DD patients who participate since 2012 in a prospective cohort study on DD course.[16]. Between 2012 and 2020, the hands of each patient have been examined every 6 months from 2012 to mid-2017 and annually since mid-2017. MHQ scores were collected between 2012 and 2014.

URAM scores were collected since 2014. The cohort and its methods are described in more detail in two previous publications.[1,16].

For this specific study we included patients who sought medical attention for DD at our outpatient clinic (clinical population). We excluded 1) hands without DD, 2) patients with no URAM or MHQ score available within the selected time frame and 3) for the MHQ analysis we excluded patients who were known to be treated during the period that the URAM was used (after we stopped collecting MHQ scores).

Ethical approval for this study was obtained from our institutional ethics review board (METc2011/397), written informed consent was obtained from all patients before the study.

Instruments

We used the Dutch version of the MHQ [11,17] between 2012 and 2014 and the Dutch version of the URAM scale[10,13] from 2014 onwards to measure self-reported hand function. We switched from the MHQ to the URAM scale because some patients with mild DD symptoms preferred a shorter questionnaire, since they reported that their disease was stable and their PROM scores were so accordingly. The MHQ contains 57 items divided over 6 subdomains: overall hand function, activities of daily living (ADL), work related activities, pain, appearance, and satisfaction with hand function. Each question can be awarded with 1 to 5 points, which are transformed to result in an overall score ranging between 0 (high disability) to 100 (no disability) for each subdomain. The total MHQ score is generated by taking the mean of the subdomain scores. The URAM scale is a 9-item questionnaire. Each question can be awarded with 0 to 5 points, with total scores ranging from 0 (no disability) to 45 (high disability). Patients completed the URAM scale for both hands separately and the MHQ according to the author's instructions (i.e., 45 questions are about both hands separately and 12 are about both hands functioning together).

Outcome measures

The primary outcome was treatment for DD during follow-up. At each visit, patients were asked if they had been treated for DD between the previous and the current study measurement or were planned for

treatment. If applicable, we recorded intervention date and type from the surgical notes of treated patients. In addition, detailed data regarding the disease course were collected, including PROM scores and passive extension deficits in each finger joint, measured in degrees by the same observer using the same finger goniometer. Of all patients their gender, age and affected hand(s) were registered.

Selection of MHQ and URAM scores

We studied whether an MHQ and URAM score measured around one year before treatment, or a change over approximately one year in MHQ and URAM score could predict treatment.

We compared patients with surgical treatment during the study period (treatment group) to clinical patients without surgical treatment (control group). Because patients at our institution are scheduled for treatment four to six weeks in advance, we excluded PROM scores measured within six weeks before treatment to avoid indication bias. Hence, we designated the treatment date minus six weeks as the index date for treated patients.

For patients in the treatment group, we selected the MHQ and URAM score measured within the time window of 15 months to six weeks prior to treatment. If multiple visits were available within this time frame, the visit closest to 12 months prior to the index date was selected as a predictor variable for treatment (**Figure 1a**). If the MHQ or URAM score on the selected study date was missing, we selected the second closest MHQ or URAM score within the time window. For the control group, the MHQ and URAM scores of a randomly selected study visit were selected.

To calculate the one-year change scores (ΔMHQ and ΔURAM), a second PROM score was needed. For treated patients, we selected the study visit closest to the index date. Then we calculated the one-year change in PROM score. For the control group, we randomly selected two study visits that were one year apart to calculate the change score (**Figure 1b**).

Statistical analysis

Patient characteristics were presented by means and standard deviations (SD) for normally distributed, continuous variables. Non-normally distributed continuous variables and ordinal variables were described by medians with interquartile ranges (IQR). There was one missing MHQ score and no

missing URAM score measured one year before treatment. For the ΔMHQ analysis, 41 (18%) hands had only one observation within our time window, so we could not calculate a change score for these patients. For the ΔURAM analysis, 46 (18%) hands had only one observation within our time window. We hypothesised that patients experiencing no functional problems because of DD were less likely to attend each follow-up visit. Hence, the missingness was probably not at random, making it impossible to impute the missing values. If the ΔMHQ/ΔURAM complete case analysis showed a good predictive ability (AUC>0.7), we planned to perform a best-case analysis to assess the influence of missing values and check the robustness of our analyses, in which we would replace each missing PROM score by the best score possible, before rerunning the analyses. In case the complete case analysis showed a poor predictive ability (AUC<0.7), we did not perform the best-case analysis.

We demonstrated the trade-off between sensitivity and specificity for specific cut-off values by drawing receiver operating characteristics (ROC)-curves (non-parametric). The MHQ/URAM scores and the ΔMHQ/ΔURAM served as test variables and the group (treatment vs. control) as state variable. Then, the area under the curve (AUC) was calculated to indicate the ability of each PROM to predict treatment. If the AUC showed at least good (>0.7) discriminative ability [18], we determined the optimal cut-off values for both MHQ/URAM and MHQ subdomain scores, by calculating 3 different metrics: 1) Youden's index (Youden, 1950), 2) Euclidean distance (Perkins and Schisterman, 2006), and 3) product of sensitivity and specificity [21]. In contrary to the positive predictive value, sensitivity and specificity are not affected by the prevalence of treated and untreated hands in the study population.[22] Therefore, we calculated the sensitivity and specificity at patient-level with 95% confidence intervals (CI) using the exact binomial method. We used the sensitivity and specificity to calculate the positive likelihood ratio (LR+) corresponding to these optimal cut-off values with 95% CIs determined by bootstrapping. The LR+ indicates how likely it is that a patient with a score equal to or lower (MHQ)/higher (URAM) than the optimal cut-off value will undergo surgical intervention in the upcoming year.

Sensitivity analysis

Patients could be affected with DD in one or both hands. For patients with DD in both hands, the PROM scores of the two hands may be correlated since some patients may generally give higher or lower scores than others. Ignoring this correlation might bias the results.[23] We determined this correlation by calculating the intra-class correlation (ICC). In the main analysis, we removed the effect of this correlation by calculating sensitivity, specificity and LR+ values on patient-level. This means that we only looked at the PROM score and treatment in a patient as a whole. So, a bilaterally affected patient with a bad PROM score in the left hand, who was treated in the right hand one year later but not in the left hand, was considered as a treated case in our main analysis. To check the robustness of our findings, we repeated the analyses on hand level, including both hands in our analysis, using multiple methods that take the correlation into account. [23]

RESULTS

Inclusion of the study population

After application of our exclusion criteria, 234 hands of 141 patients remained for the MHQ analysis and 251 hands of 145 patients for the URAM analysis (**Figure 2)**. Patients were treated with limited fasciectomy, percutaneous needle fasciotomy, dermofasciectomy or clostridium collagenase histolyticum.

Population characteristics

For the MHQ analysis (n = 141), the treatment group comprised 20 (9%) hands and the control group comprised 214 (91%) hands. The median age was 64.0 (IQR 58.0-70.0) years, and 98 (70%) patients were male. For the URAM analysis (n = 145), the treatment group comprised 58 (23%) hands and the control group comprised 193 (77%) hands. The median age was 66.0 (IQR 61.0-73.0) years, and 104 (72%) patients were male. The median MHQ score measured one year before treatment was 76.5 (IQR 54.3 - 85.3) for treated hands and 91.9 (IQR 81.8 - 99.2) for untreated hands. For the URAM, the median score 6.5 (IQR 2.8 - 14.0) for treated hands and 1.0 (IQR 0.0 - 5.0) for untreated hands (**Table 1)**. The median scores of the two selected MHQ and URAM across time seemed to be worse for patients

who got treatment compared with controls, although the variability in the measurements was high (**Figure 3**).

MHQ score

The median interval between the date of the selected PROM score in the study interval and treatment was 8.5 months (IQR 5.7-10.9) for the MHQ. For the MHQ score, the ROC-curve showed an AUC of 0.80 (95% CI 0.71–0.88) **(Figure 4a**). For the total MHQ score and four MHQ subdomains (overall hand function, ADL, work and satisfaction), the Youden's index, Euclidian distance and product method resulted in two or three different optimal cut-off values and corresponding LR+s. For the other two MHQ subdomains (pain and appearance), the three metrics resulted in the same cut-off value and corresponding LR+s. (**Table 2).** The results of the sensitivity analyses (hand-level) were very similar to the results on patient level.

URAM score

The median interval between the date of the selected PROM score in the study interval and treatment was 11.5 months (IQR 8.4-13.2). The ROC-curve showed an area under the curve (AUC) of 0.75 (95% CI 0.68-0.82) (**Figure 4b**). The Youden's index, the Euclidean distance and the product method resulted in two different optimal cut-off value and corresponding LR+ (**Table 3).** Again, the sensitivity analyses produced very similar results.

ΔMHQ and ΔURAM

The median interval between the two MHQ scores was 5.8 months (IQR 4.2-7.0) for treated patients and 12.0 months (IQR 11.7-12.2) for untreated patients. For the URAM, the median interval was 6.7 months (IQR 5.7-10.6) for treated patients and 12.0 months (IQR 11.1-12.6) for untreated patients). The one-year ΔMHQ score showed an AUC of 0.54 (95% CI 0.38–0.71) and the one-year ΔURAM score showed an AUC of 0.57 (95% CI 0.44-0.69) **(Supplementary Figure 1)**. Because an AUC of <0.70 indicates a poor predictive ability, we did not calculate the LR+ for the Δ MHQ and ΔURAM scores and we did not perform a best-case analysis.

DISCUSSION

Our results show that both the MHQ and URAM score measured around one year before treatment is predictive of treatment for DD. The one-year change score appeared to be a poor predictor for treatment.

We found that both the MHQ and URAM scale can be used to predict treatment with a good discriminative ability (AUC 0.80 and 0.75, respectively).[18] Because of the high sensitivity of both the MHQ and URAM scores for predicting treatment, the chance of missing patients in need for treatment is small. In a previous study, MHQ and URAM scores appeared to be predictive of an increase of total passive extension deficit (TPED) on group level, but not on individual level.[10] It is likely that surgery is more predictable than TPED increase, because (change in) TPED is subject to margins of error.[24] To our knowledge, no previous studies in the field of hand surgery have been conducted with the aim to predict surgery. However, multiple studies within other surgical fields did report on the use of PROMs to predict postoperative outcomes. Preoperatively assessed PROM scores could successfully predict long and short term post-operative improvement on physical function and pain after foot, ankle and shoulder surgery with similar discriminative abilities (AUC>0.70).[25–29] Although the preoperative PROM scores in these studies were measured shortly before surgery and therefore cannot be compared on a one-to-one basis with our cohort study, the results are in line with our conclusion that PROM scores can be used to predict clinical outcomes.

In this study, the MHQ subdomain ‘satisfaction with hand function’ had the highest predictive ability for DD treatment (AUC 0.82), and the ‘work’ domain had the lowest predictive ability (AUC 0.66) compared to the remaining four subscales. This is in concordance with a recent study on the effect of treatment of Dupuytren’s disease on the different MHQ domains three months after surgery. [30] The authors found that a decrease in extension deficit mainly improved the ‘appearance of the hand’ and the ‘satisfaction with hand function’ subscales, while the ‘work’ subscale showed no significant treatment effect.[29] In our study, the appearance of the hand was less predictive than most other domains. Potentially, appearance is associated with improvement of extension deficits shortly after DD treatment

but is less sensitive as a predictor for DD treatment. The good predictive ability of the 'pain' subdomain (AUC 0.71) is remarkable, since it has been reported that pain is rare and mostly mild in DD patients.[31] It could be that pain is relatively more common in patients with active disease, and less present in patients with stable disease, but no evidence is available yet to support this theory.

We hypothesised that not only the self-reported hand function measured at a specific time point but also a decline in self-reported hand function could be predictive for DD treatment, since inter-individual differences at baseline would be eliminated using a change score. Our results however demonstrated that the one-year change in MHQ and URAM scores has no predictive power (AUC of 0.52 and 0.56). We are not aware of other studies that used a change in preoperative measured PROM scores to predict treatment. In theory, a one-year interval between completing two MHQ or URAM scales might be too short to identify a clinical important change in hand function, especially in patients with early, slowly progressing DD [32]. On the other hand, applying a longer interval would risk missing patients with aggressive disease with rapid disease progression.[16] The low predictive ability of the change score could also be explained by the large intra-individual variance in self-reported hand function that we observed in our study. This means that the functional disabilities reported by a patient varied over time, without having undergone any treatment. Instead of using the change in PROM score over time, it may be appropriate to monitor patients every 12 months to determine whether the MHQ or URAM score has exceeded the optimal cut-off value.

One of the strengths of this study is that we included clinical DD patients with various disease severities. Because of the unique longitudinal nature of our data, these patients assessed their self-reported hand function regularly, often long before treatment would be indicated. Hence, we minimised the risk of selection and indication bias. Secondly, patient reported hand function appears to be predictive of surgery, despite the many other factors that play a role in this decision, such as age and occupational status. Moreover, we tested our hypothesis using two different PROMs in one study population. This substantiates our conclusion that PROMs can be used to predict treatment for DD.

Our study has some limitations. First, we were not able to set a single cut-off value for both PROMs. There is no single method for determining optimal cut-off values in ROC curves. The choice between these methods depends on variability of test results in diseased and non-diseased subjects and the desired sensitivity that is most clinically relevant.[33] Because the desired sensitivity for the aim of this study is yet unknown, we used three different metrics: the Youden-index, the Euclidean distance and the product index. Since these metrices all maximise the overall correct classification rate and assign equal weight to the sensitivity and the specificity,[34] we consider all three methods appropriate for this experimental study. For the total MHQ score and four MHQ subdomains, the use of these three metrics resulted in different optimal cut-off values. Before we can set a single cut-off value for both PROMs, a budget-impact analysis should clarify the desired sensitivity of the test and indicate which cut-off value would be best to use in clinical practice. Secondly, we did not stratify between patients with primary disease and recurrent disease. Patients do not always regain optimal self-reported hand function after surgery.[30] If we would have analysed these patients separately, we might have found lower optimal cut-off values for patients with recurrent disease. However, we decided to include all patients as one group to maintain a large sample size and because it is desirable in clinical practice to apply remote patient monitoring in both primary and recurrent disease.

Our study shows that MHQ and URAM scores are predictive of DD treatment. If future budget-impact studies show that telemonitoring of DD patients with PROMs also leads to cost reduction, web-based PROMs could optimise patient care and treatment effectiveness of Dupuytren's disease.

Acknowledgements

We thank Katalin Tamasi, PhD, for her statistical expertise in handling clustered data and performing the sensitivity analyses.

REFERENCES

1. Broekstra DC, Lanting R, Werker PMN, van den Heuvel ER. Disease Course of Primary Dupuytren Disease: 5-Year Results of a Prospective Cohort Study. *Plast Reconstr Surg*. 2022 Jun;149(6):1371–8.
2. Molenkamp S, Schouten TAM, Broekstra DC, Werker PMN, Moolenburgh JD. Early Postoperative Results of Percutaneous Needle Fasciotomy in 451 Patients with Dupuytren Disease. *Plast Reconstr Surg*. 2017 Jun;139(6):1415–21.
3. Stepić N, Končar J, Rajović M. The influence of Dupuytren's disease fingers contracture degree on surgical treatment outcome. *Vojnosanit Pregl*. 2017 Jan;74(1):19–23.
4. van Rijssen AL, Gerbrandy FSJ, ter Linden H, Klip H, Werker PMN. A comparison of the direct outcomes of percutaneous needle fasciotomy and limited fasciectomy for Dupuytren's disease: a 6-week follow-up study. *J Hand Surg Am*. 2006;31(5):717–25.
5. de Farias FACD, Dagostini CM, Bicca YDA, Falavigna VF, Falavigna A. Remote patient monitoring: A systematic review. *Telemed J E Health*. 2020 May 1;26(5):576–83.
6. Wiegel J, Seppen B, Leeden M van der, et al. Adherence to Telemonitoring by Electronic Patient-Reported Outcome Measures in Patients with Chronic Diseases: A Systematic Review. *Int J Environ Res Public Health*. 2021 Oct 1;18(19):10161.
7. de Jong M, van der Meulen-de Jong A, Romberg-Camps M, et al. Telemedicine for management of inflammatory bowel disease (myIBDcoach): a pragmatic, multicentre, randomised controlled trial. *Lancet*. 2017 Sep 2;390(10098):959–68.
8. Solomon D, Lu F, Xu C, et al. Patient adherence with a smartphone app for disease monitoring in rheumatoid arthritis. *Ann Rheum Dis*. 2019 Jun 1;78(Suppl 2):350–1.
9. Vigneau C, Choukroun G, Isnard-Bagnis C, et al. A patient centered study using an electronic choice-based conjoint analysis (ePRO) to assess real world preferences regarding erythropoiesis stimulating agent to treat anaemia in chronic kidney disease (PERCEPOLIS study). *Nephrol Ther*. 2019 Jun 1;15(3):152–61.

10. Broekstra DC, van den Heuvel ER, Lanting R, Werker PMN. Measurement properties of the Dutch Unité Rhumatologique des Affections de la Main and its ability to measure change due to Dupuytren's disease progression compared with the Michigan Hand outcomes Questionnaire. *Journal of Hand Surgery: European Volume*. 2018 Oct 1;43(8):855–63.

11. Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. *J Hand Surg Am*. 1998 Jul 1;23(4):575–87.

12. Marks M. Which patient-reported outcomes shall we use in hand surgery? *J Hand Surg Eur Vol*. 2020 Jan 1;45(1):5–11.

13. Beaudreuil J, Allard A, Zerkak D, et al. Unité Rhumatologique des Affections de la Main (URAM) scale: development and validation of a tool to assess Dupuytren's disease-specific disability. *Arthritis Care Res (Hoboken)*. 2011 Oct;63(10):1448–55.

14. Bradet-Levesque I, Audet J, Roy J-S, Flamand VH. Measuring functional outcome in Dupuytren's disease: A systematic review of patient-reported outcome measures. *J Hand Ther*. 2021 Apr;(in press)(0).

15. Scheibler AG, Marks M, Hensler S, Herren DB, Calcagni M. Factors predicting the 1-year outcome of collagenase treatment for Dupuytren's disease. *Arch Orthop Trauma Surg*. 2019 Apr 1;139(4):583–8.

16. Lanting R, van den Heuvel ER, Werker PMN. Clusters in Short-term Disease Course in Participants with Primary Dupuytren Disease. *J Hand Surg Am*. 2016;41(3):354–61.

17. Schoneveld K, Wittink H, Takken T. The Dutch language version of the Michigan Hand Outcomes Questionnaire: valdiation in patients with Dupuytren's disease. 2009;119:161–9.

18. Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. 2nd Ed. New York; 2000. p. 160-164.

19. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950 Jan;3(1):32–5.

20. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol*. 2006 Apr;163(7):670–5.

21. Liu X. Classification accuracy and cut point selection. *Stat Med*. 2012 Oct 15;31(23):2676–86.

22. Ranganathan P, Aggarwal R. Understanding the properties of diagnostic tests - Part 2: Likelihood ratios. *Perspect Clin Res*. 2018;9(2):99.

23. Genders TSS, Spronk S, Stijnen T, et al. Methods for calculating sensitivity and specificity of clustered data: A tutorial. *Radiology*. 2012;265(3):910–6.

24. Broekstra DC, Lanting R, Werker PMN, van den Heuvel ER. Intra- and inter-observer agreement on diagnosis of Dupuytren disease, measurements of severity of contracture, and disease extent. *Man Ther*. 2015 Aug;20(4):580–6.

25. Waly FJ, Yeo EMN, Wing KJ, et al. Relationship of Preoperative Patient-Reported Outcome Measures (PROMs) to Postoperative Success in End-Stage Ankle Arthritis. *Foot Ankle Int*. 2020 Mar 1;41(3):253–8.

26. Ho B, Houck JR, Flemister AS, et al. Preoperative PROMIS Scores Predict Postoperative Success in Foot and Ankle Patients. *Foot Ankle Int*. 2016 Sep 1;37(9):911–8.

27. Chen RE, Papuga MO, Nicandri GT, Miller RJ, Voloshin I. Preoperative Patient-Reported Outcomes Measurement Information System (PROMIS) scores predict postoperative outcome in total shoulder arthroplasty patients. *J Shoulder Elbow Surg*. 2019 Mar 1;28(3):547–54.

28. Anderson MR, Houck JR, Saltzman CL, et al. Validation and Generalizability of Preoperative PROMIS Scores to Predict Postoperative Success in Foot and Ankle Patients. *Foot Ankle Int*. 2018 Jul 1;39(7):763–70.

29. Giesinger JM, Hamilton DF, Jost B, Behrend H, Giesinger K. WOMAC, EQ-5D and Knee Society Score Thresholds for Treatment Success After Total Knee Arthroplasty. *J Arthroplasty*. 2015 Dec 1;30(12):2154–8.

30. Poelstra R, van Kooij YE, van der Oest MJW, et al. Patient's satisfaction beyond hand function in Dupuytren's disease: analysis of 1106 patients. *J Hand Surg Eur*. 2020 Mar;45(3):280–5.

31. Vandecasteele L, Degreef I. Pain in Dupuytren's disease. *Acta Orthop Belg*. 2020;86:555–62.

32. van den Berge BA, Werker PMN, Broekstra DC. Limited progression of subclinical Dupuytren's disease. *Bone Joint J*. 2021 Apr;103-B(4):704–10.

33. Hajian-Tilaki K. The choice of methods in determining the optimal cut-off value for quantitative diagnostic test evaluation. *Stat Methods Med Res*. 2018 Aug 1;27(8):2374–83.

34. Unal I. Defining an Optimal Cut-Point Value in ROC Analysis: An Alternative Approach. *Comput Math Methods Med*. 2017;2017.

FIGURE LEGENDS

Figure 1a. Selection of the total MHQ or URAM score. We selected the score measured closest to 12 months before treatment within the time interval. For the control group, we randomly selected a study visit.

Figure 1b. Selection of the one-year ΔMHQ and ΔURAM score. We selected a second visit being closest to 6 weeks before treatment. For the control group, we randomly selected two study visits with a one-year interval.

Figure 2. Flowchart of the inclusion process of patients and hands. Patients could be included with either one or both hands. Hands that met the exclusion criteria were excluded. Only when both hands of a patient were excluded, the patient was excluded, explaining why more hands than patients were excluded at each step.

Figure 3 Boxplots of the two selected MHQ **(a)** and URAM **(b)** scores, with the first visit being closest to one year before treatment and second visit being closest to 6 weeks before treatment.

A high URAM score represents high disability. A high MHQ score represents low disability.

Figure 4a. ROC curve for the MHQ score to predict treatment

Figure 4b. ROC curve for the URAM score to predict treatment

Supplementary figure 1a. ROC curve for the one-year change of MHQ score to predict treatment

Supplementary figure 1b. ROC curve for the one-year change of URAM score to predict treatment

Table 1: Descriptive statistics of the MHQ and URAM scores of treated and untreated hands measured one year before treatment.

	Treated	Untreated
MHQ		
n	20	214
Total score	76.5 (54.3 - 85.3)	91.9 (81.8 - 99.2)
Overall HF	62.5 (50.0 - 78.8)	80.0 (70.0 - 100.0)
ADL	82.5 (65.0 - 100.0)	100.0 (90.0 - 100.0)
Pain	75.0 (55.0 - 100.0)	95.0 (75.0 - 100.0)
Work	85.0 (60.0 - 100.0)	100.0 (85.0 - 100.0)
Aesthetics	81.3 (54.7 - 92.2)	100.0 (87.5 - 100.0)
Satisfaction	62.5 (27.1 - 85.4)	95.8 (78.1 - 100.0)
URAM		
n	58	193
Total score	6.5 (2.8 - 14.0)	1.0 (0.0 - 5.0)

HF = hand fuction; ADL = activities of daily living;

PROM scores are presented as medians (interquartile range).

Table 2. Optimal cut-off values for MHQ score, determined using three different methods, and the corresponding sensitivity, specificity and positive likelihood ratio.

	AUC (95% CI)	Cut-off	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)
MHQ total	**0.80 (0.71-0.88)**				
Youden		91.2	0.95 (0.74-1.00)	0.48 (0.39-0.58)	1.83 (1.47-2.23)
Euclidean & product		83.9	0.74 (0.49-0.91)	0.68 (0.59-0.76)	2.30 (1.52-3.31)
Overall HF	**0.75 (0.65-0.86)**				
Youden		87.5	0.95 (0.74-1.00)	0.35 (0.27-0.44)	1.46 (1.20-1.70)
Euclidean & product		72.5	0.63 (0.38-0.84)	0.60 (0.51-0.69)	1.57 (0.95-2.26)
ADL	**0.73 (0.60-0.86)**				
Youden		72.5	0.37 (0.16-0.62)	0.90 (0.83-0.95)	3.75 (1.42-8.52)
Euclidean & product		92.5	0.63 (0.38-0.84)	0.66 (0.57-0.75)	1.88 (1.13-2.79)
Pain	**0.71 (0.59-0.83)**				
Youden, Euclidian & product		82.5	0.74 (0.49-0.91)	0.71 (0.62-0.79)	2.57 (1.67-3.76)
Work	**0.66 (0.53-0.80)**				
Youden		82.5	0.53 (0.29-0.76)	0.82 (0.74-0.88)	2.92 (1.49-5.10)
Euclidean & product		97.5	0.63 (0.38-0.84)	0.68 (0.59-0.76)	1.98 (1.19-2.96)
Appearance	**0.73 (0.62-0.85)**				
Youden, Euclidian & product		90.6	0.79 (0.54-0.94)	0.65 (0.56-0.73)	2.24 (1.55-3.10)
Satisfaction	**0.82 (0.75-0.89)**				
Youden & product		93.8	1.00 (0.82-1.00)	0.43 (0.34-0.53)	1.77 (1.42-2.05)
Euclidean		81.3	0.74 (0.49-0.91)	0.62 (0.53-0.71)	1.95 (1.30-2.72)

HF = hand function; ADL = activities of daily living.

Table 3. Optimal cut-off values for URAM score, determined using three different methods, and the corresponding sensitivity, specificity and positive likelihood ratio.

	AUC (95% CI)	Cut-off	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)
Total score	**0.75 (0.68-0.82)**				
Youden & product		2.5	0.78 (0.64-0.88)	0.44 (0.34-0.55)	1.40 (1.10-1.77)
Euclidean		3.5	0.70 (0.55-0.82)	0.51 (0.41-0.60)	1.41 (1.07-1.86)

Figure 1

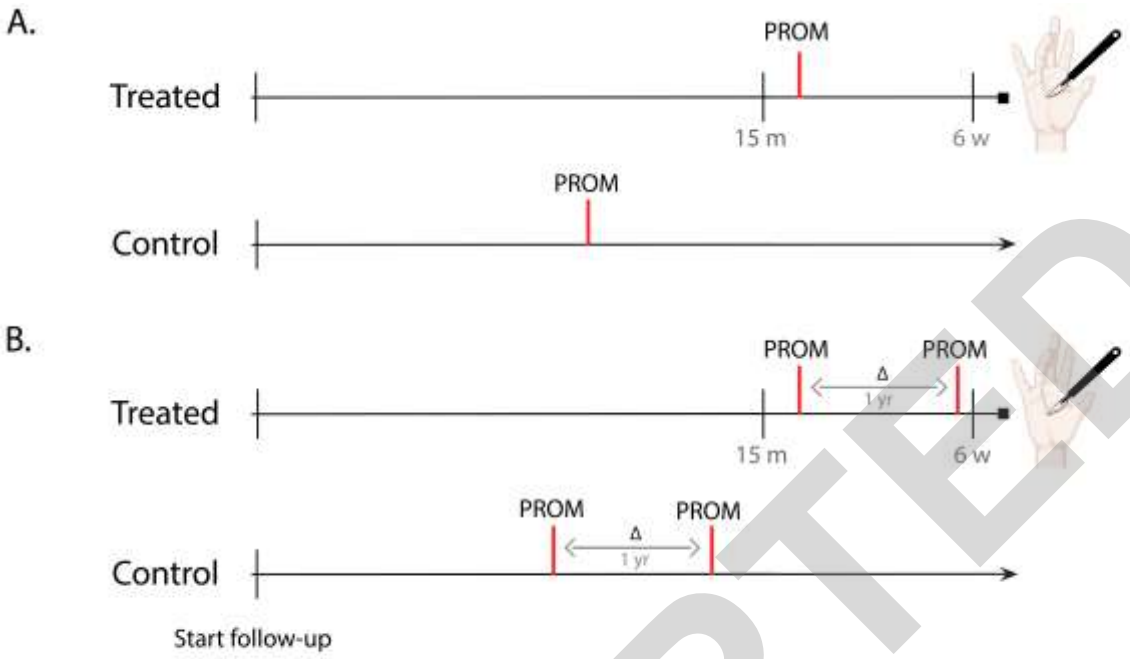


Figure 2

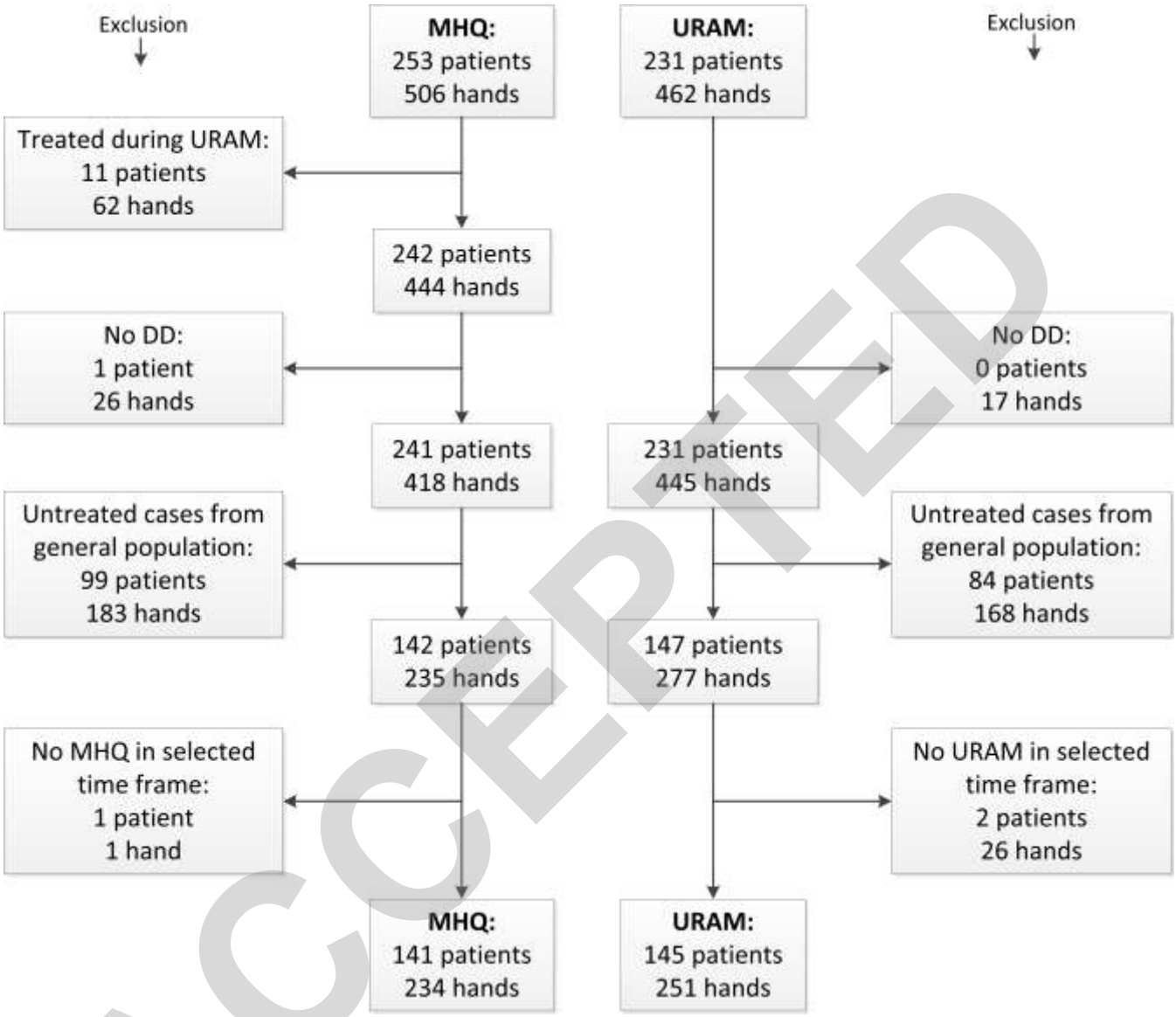


Figure 3a

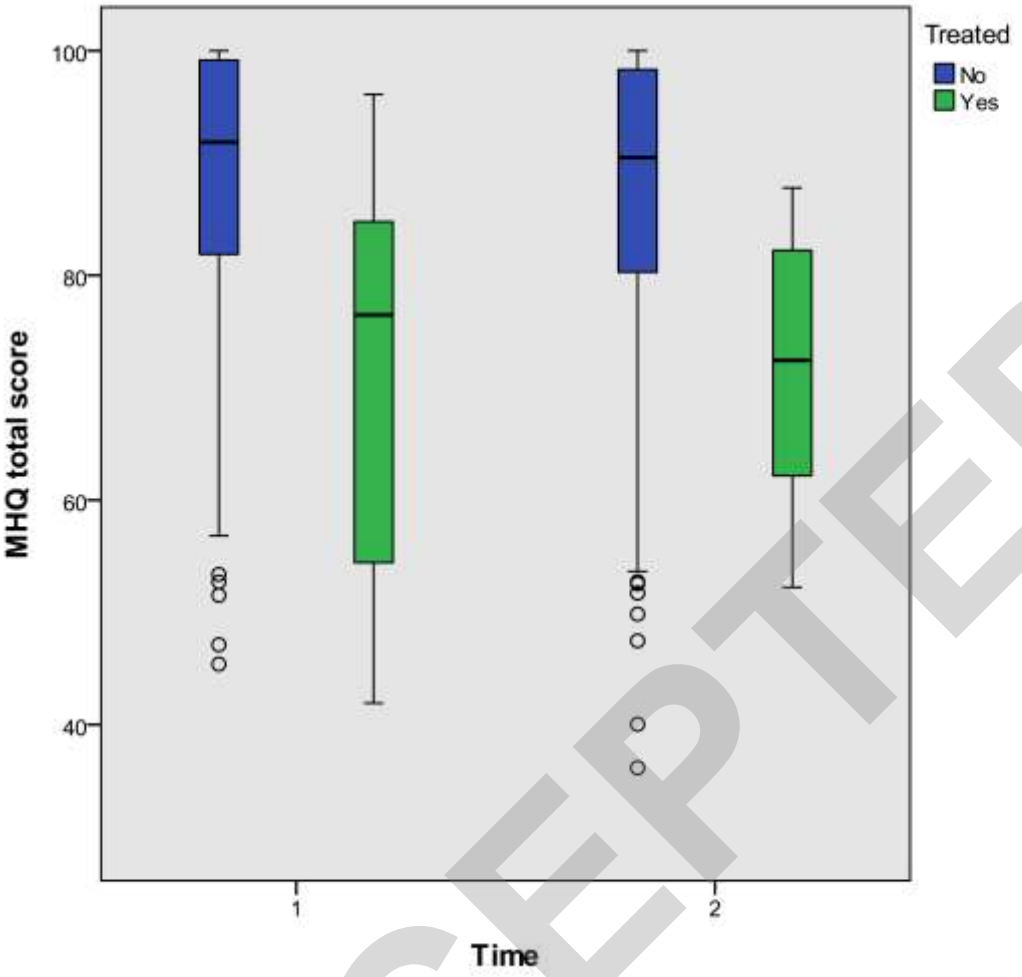


Figure 3b

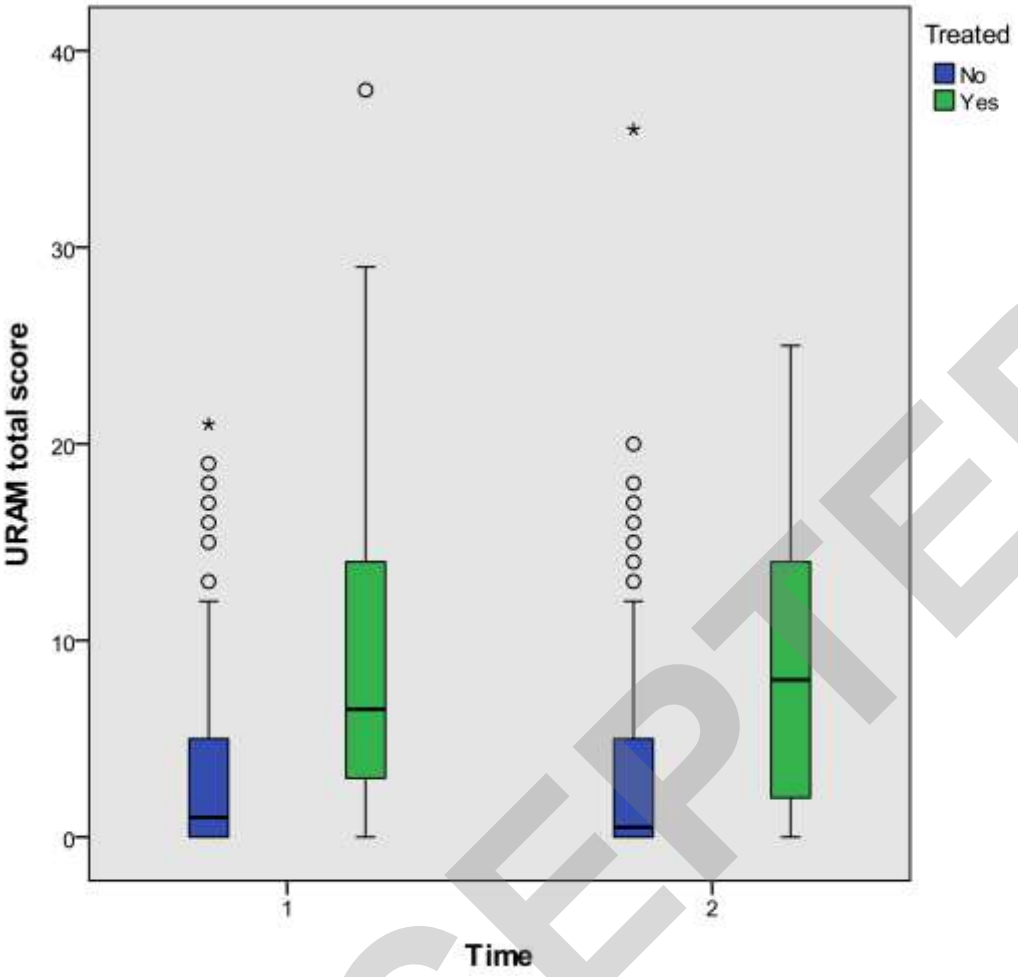


Figure 4a

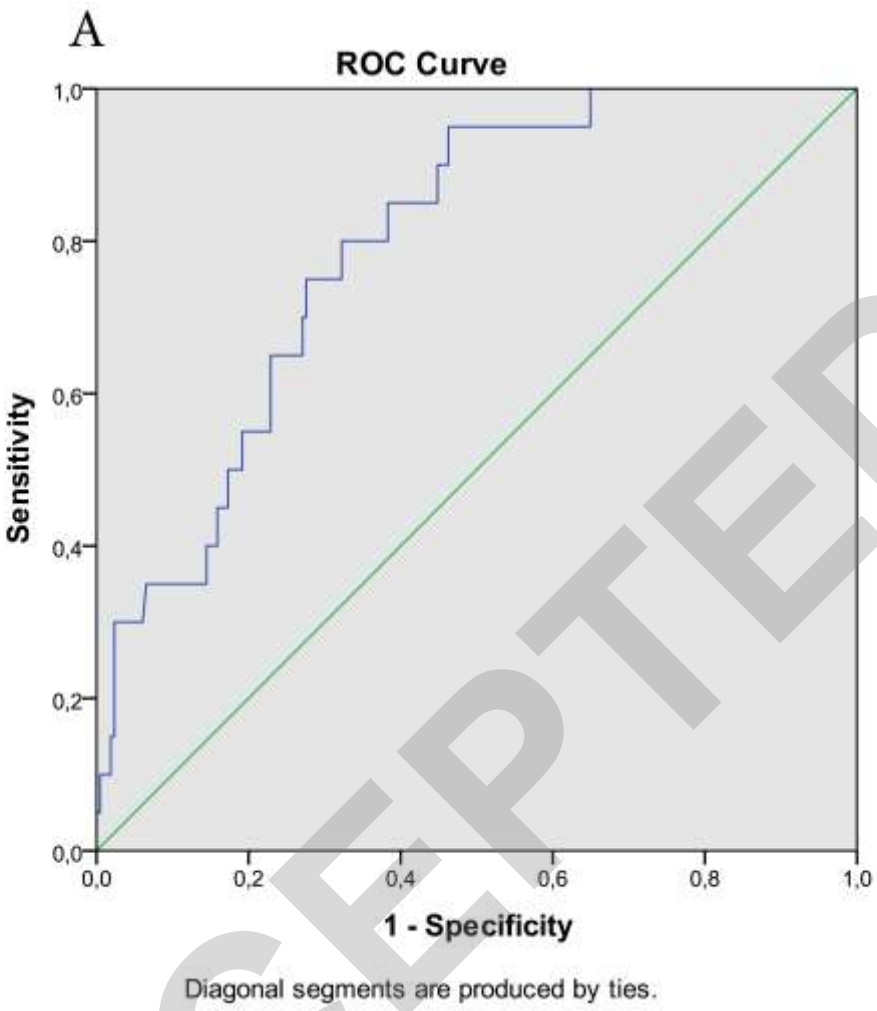
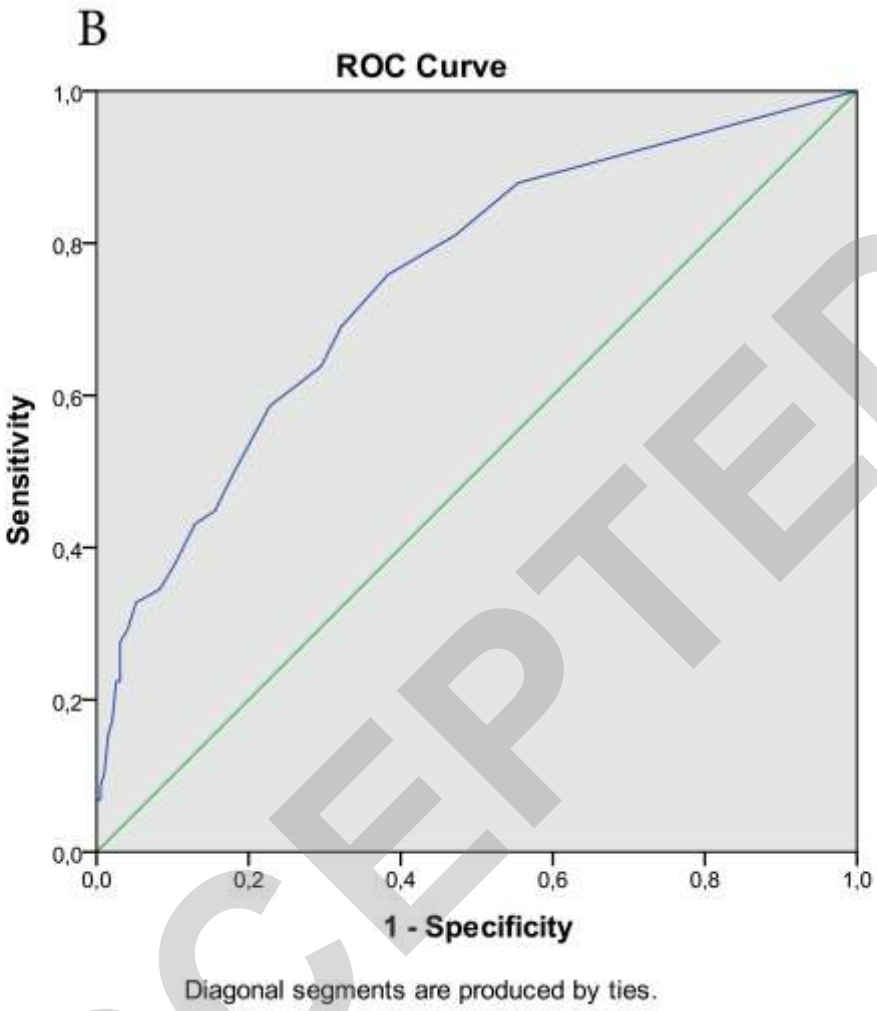
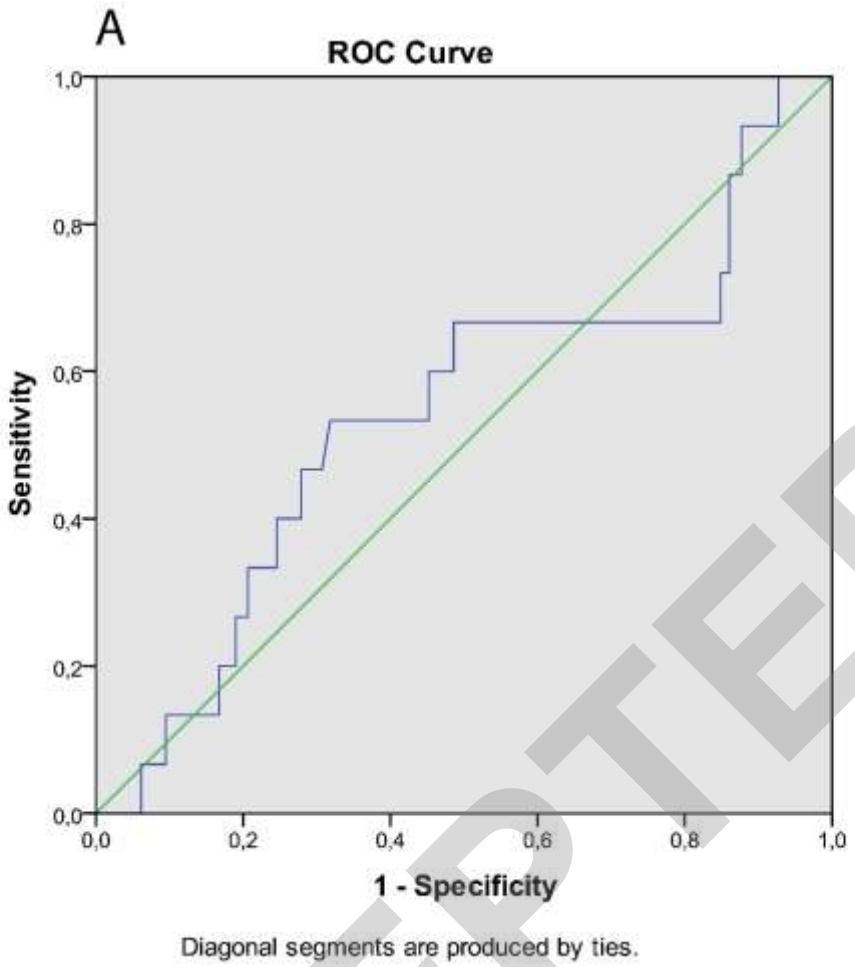
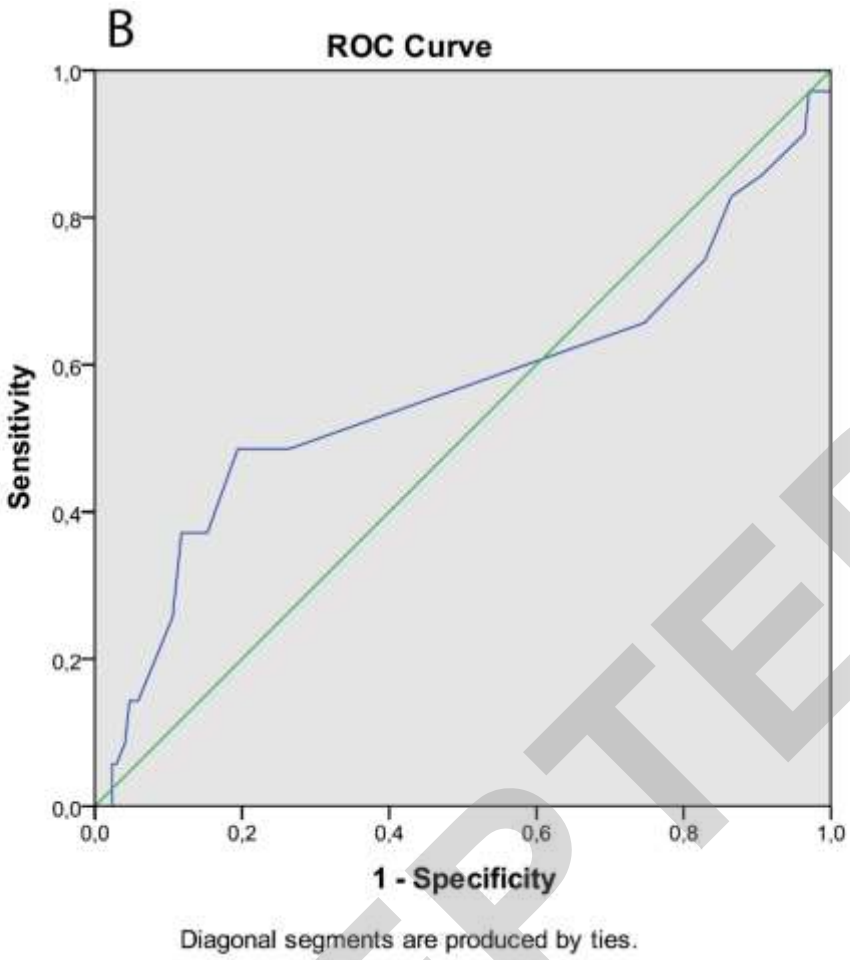


Figure 4b





Supplementary figure 1a. ROC curve for the one-year change of MHQ score to predict treatment



Supplementary figure 1b. ROC curve for the one-year change of URAM score to predict treatment