Supplemental Digital Content 1. Statistical methodology

The statistical analysis method used a generalized linear mixed model on the frequencies with a binomial distribution and a logit link function. The model treats study as a random effect on the log odds values. The analysis was conducted with procedure NLMIXED of SAS version 9.4, where the variance component for study was modeled in the logarithmic scale to avoid boundary issues. The association of Dupuytren and the diseases was estimated with a pooled log odds parameter over all studies.

The association between DD and DM was adjusted for the potential confounding effect of age. In case age was reported in categories, the mean age was estimated using the midpoints of each category. The association between DD and liver disease was adjusted for the proportion of men in each group to control for the potential confounding effect of sex. When a study matched participants on sex while the sex distribution was not reported for subgroups separately, the missing information was imputed using a linear regression model. The variables study ID, group (disease or control), number of Dupuytren cases, and number of total participants were used as predictors, to impute the proportion of men. Twenty imputations were performed, with the maximum number of iterations set at 10. The analysis was conducted for each imputed data and Rubin's approach was applied to the model parameters using procedure MIANALYZE of SAS version 9.4. For the association between DD and epilepsy no adjustment for age or sex was applied. We chose for this methodology, as we assumed that there were no confounders for the association between epilepsy and DD.

Heterogeneity was calculated with the intraclass correlation coefficient (ICC):

$$ICC = \frac{\sigma^2_{random}}{\sigma^2_{random} + (\frac{\pi^2}{3})}$$

The larger the ICC value, the larger the heterogeneity. This value can be interpreted as a measure of consistency.