Comparison of two natural history studies as part of the HD-IH consortium: a use case for historical external control group matching in clinical trials

M. Papoutsi[1,2], J.A. Mills[3], E.C. Gantman[4], C. Sampaio[4], J.D. Long[3], R. Wolz[1,4]

[1]IXICO plc, London UK; [2]University College London, London, UK; [3]University of Iowa, IA, USA; [4]CHDI, New York, NY, USA; [5]Imperial College London, London UK

External control groups can support regulatory decision making in clinical trials, especially rare diseases. In Huntington's disease (HD) there is a wealth of historical data, but matching historical data to clinical trials is challenging.

Here, we show how historical data from two natural history studies, TRACK-HD and PREDICT-HD, have sufficiently similar sub-populations to allow their merging and use as an external comparator group. To evidence this we selected matched samples from the two studies using propensity scores and compared groups using change over time in caudate, putamen, whole brain and lateral ventricles volume, and composite Unified HD Rating Scale (cUHDRS), Symbol Digit Modalities Test (SDMT), Total Motor Score (TMS) and Total Functional Capacity (TFC).

METHODS

Using data from the HD Imaging Harmonization (HD-IH) consortium we selected 87 participants from TRACK-HD using criteria: 25-65 years old, CAG repeat length (CAG) 40-50 and HD-ISS Stage 2 or 3 and TFC > 10 (Tabrizi et al, 2022). These criteria are similar to a putative clinical trial.

Participants also had to have a 2-year follow-up MRI scan, because in PREDICT-HD most scans occurred bi-annually. Using very different time-intervals is not recommended when matching studies, even if annualized. This is because of differences in sensitivity across time either due to algorithm bias in the case of volumetry or practice effects in the case of clinical scores.

For matching, we first selected PREDICT-HD participants with a 2-year follow-up and 40-50 CAG. We then applied optimal propensity score matching using logistic regression (Lynch et al. 2024) using the baseline prognostic index for HD (PIN-HD) score, sex, age, CAG, and their interaction. Matching was exact for PIN-HD which resulted in highly balanced cohorts (distance standardized mean difference = 0.01, variance ratio = 1.03). This resulted in the selection a subset of 76 matched cases from both studies.

Variables	TRACK-HD	PREDICT-HD
N	76	76
Years on Study (time from baseline) – Mean (SD)	1.99 (0.11)	2.03 (0.11)
Age – Mean (SD)	45.2 (9.4)	45.5 (9.7)
Sex (%F)	56.6%	57.9%
CAG – Median (min-max)	43 (40 – 50)	43 (40 – 47)
PIN-HD – Mean (SD)	1.06 (0.92)	1.06 (0.92)

To compare the matched groups, we fit linear mixed models with random intercepts with annualized change in regional volume (normalized for head size), TMS, SDMT, TFC or cUHDRS as outcomes. Models were adjusted for time (1 year), group, baseline outcome value, and their interaction with time. TFC, SDMT and cUHDRS models were also adjusted for education (high vs low) and its interaction with time. **Separate models were fit for each outcome**. The full model is depicted below.

$$Annualized\ outcome\ change = 1 + time \times (baseline\ outcome + Study + [education]) + (1|\ participants)$$

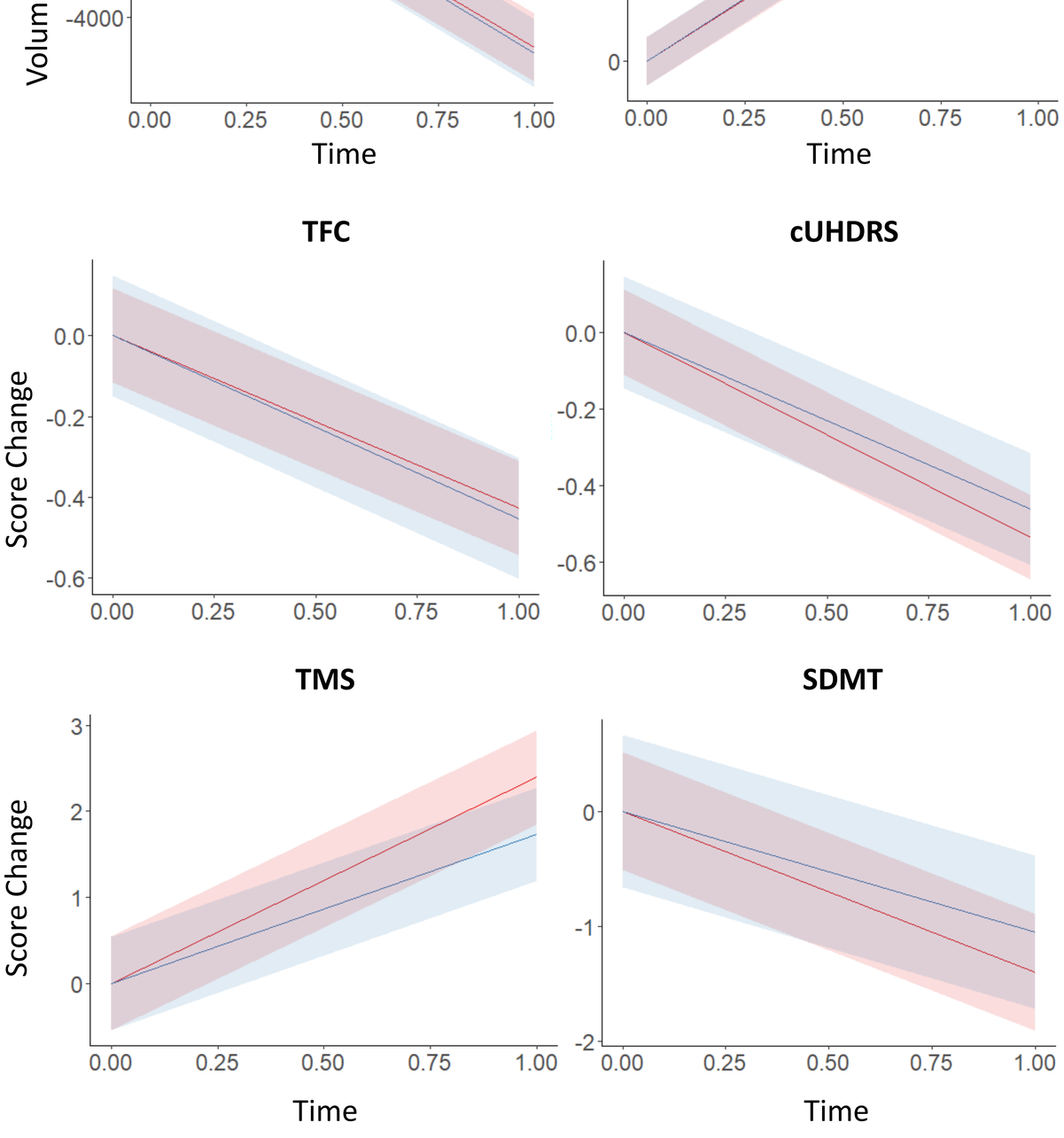
LMM RESULTS: Difference in slope between studies

The table below reports the LMM statistical results for the interaction between change and study (TRACK-HD vs PREDICT-HD) on the annualized change in each of the eight outcome measures. None of the comparisons was significant, therefore p-values were not corrected for multiple comparisons.

Outcome	Contrast Estimate (SE)	T-value	P-value
Annualized Caudate Nucleus Volume Change	2.01 (15.42)	0.13	0.896
Annualized Putamen Change	42.85 (23.32)	1.84	0.068
Annualized Whole Brain Change	124.22 (717.67)	0.17	0.863
Annualized Lateral Ventricles Change	-24.37 (119.68)	-0.20	0.839
Annualized TFC Change	0.03 (0.12)	0.23	0.822
Annualized cUHDRS Change	-0.07 (0.12)	-0.62	0.535
Annualized TMS Change	0.66 (0.56)	1.19	0.237
Annualized SDMT Change	-0.35 (0.55)	-0.64	0.524

RESULTS

Fitted slopes showing the Study by change interaction (95% CI)



SUMMARY

For HD-ISS Stage 2 and early Stage 3 participants there was no significant effect of study on change for any of the eight outcomes measured here (all p > 0.05 uncorrected). The two matched groups had similar trajectories over 2 years.

Our work therefore shows that data from the two studies are sufficiently similar to be merged and considered as an external comparator group in clinical trials.

References: • Tabrizi et al., 2022, Lancet Neurology 12 (7) • Lynch et al., Ann Clin Transl Neurol 11 (1)