Muscle strength measures as endpoints for clinical trials in muscle disorders

Challenges in Drug Development for Muscle Disease Workshop
Clinically Significant Endpoints in Muscle Disease
August 4-5, 2005

Diana M. Escolar, MD
Associate Prof. OF Neurology and Pediatrics
Children's National Medical Center
George Washington University

The issues with the various approaches to measure muscle strength

- Reliability
  - Can the measure be replicated with small variability? Inter-rater? Intra-rater? Decreasing the measure error = smaller sample size needed
- Sensitivity
  - Able to detect small or large intervention effects
  - Different interventions might require more or less sensitive measures
- Clinical meaningfulness
  - Is change in muscle strength a clinically relevant outcome in muscle diseases?
  - If so, to what extend? What is a clinically significant change in muscle strength For the individual patient (patient effect)? For the population measured (effect size)?? Is it different for different muscle diseases?
  - Correlation with patient reported and/or physician reported outcomes
- Validity
  - Is muscle strength a valid trial endpoint that may serve as surrogate for clinical benefit?
    - the surrogate endpoint must correlate with the clinical endpoint and fully capture the net effect of treatment on the clinical outcome
    - a convincing surrogate endpoint must be biologically relevant as well as epidemiologically coherent
- Feasibility

Terminology
(From the Merriam-Webster Dictionary)

- Strength
  - 1 : the quality or state of being strong : capacity for exertion or endurance
  - 2 : power to resist force: solidity, toughness
  - Antonym: weakness
- Fatigue
  - 2 a : weariness or exhaustion from labor, exertion, or stress b : the temporary loss of power to respond induced in a sensory receptor or motor end organ by continued stimulation
  - 3 : the tendency of a material to break under repeated stress
  - Antonym: energy
- Endurance
  - **1 : Permanence, Duration**
  - **2 :** the ability to withstand hardship or adversity; *especially* **:** the ability to sustain a prolonged stressful effort or activity <a marathon runner's *endurance*>
    Valid measure: 6 minute Walking test

Types of muscle movement

- 1. Isotonic, also called dynamic, that means equal tone, or constant load throughout the range of motion (ROM). Its characteristics are:
  - the length of the muscle changes
  - the muscle must overcome the resistance
  - the same load is applied throughout the range of motion (ROM)
  - the speed of the motion may change throughout the ROM.
- 2. Isometric, also called static, and means equal muscle length, because there is no external motion. Its characteristics are:
  - the length of the muscle does not change
  - the muscle cannot overcome the resistance.
- 3. Isokinetic, which means equal velocity of motion. Its characteristics are:
  - the length of the muscle changes
  - the velocity of the motion does not change throughout the ROM
  - the force produced changes throughout the ROM
  - requires expensive equipment for the testing process.

Measuring muscle strength with isometric technique

- **Manual muscle testing**

  The Break Test

  – Manual resistance is applied to a limb or

  other body part at the point in the ROM where

  the muscle is most challenged.

  – The term resistance denotes the force that acts in opposition to the contracting muscles

  – The break test requires the subject to hold the position and not let the examiner break the hold

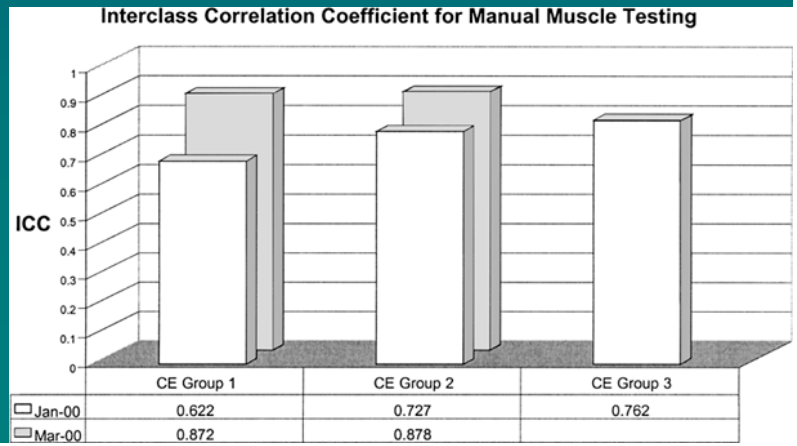
MMT Grading Schemes

Daniels	Kendall	MRC
Normal	100%	5
Good	80%	4
Fair	50%	3
Poor	20%	2
Trace	5%	1
Zero	0%	0

MMT score: Reliable? Duchenne

- **CIDD**
  - **4 CEs** FAIR for individual muscle groups "The reliability of grades for individual muscle groups ranged from .65 to .93, with the proximal muscles having the higher reliability values. The reliability of individual muscle strength grades ranged from .80 to .99, with those in the gravity-eliminated range scoring the highest. We conclude the MMT grades are reliable for assessing muscle strength in boys with DMD when consecutive evaluations are performed by the same physical therapist" … Florence et al, Phys Ther 1992
  - 4 CEs ICC intra-rater 0.95 and intra-rater 0.90 (learning curve)
- **CINRG** 12 CEs

Clinical evaluator reliability for quantitative and manual muscle testing measures of strength in children. D. Escolar, E. Hnericson, J. Mayhew,, J. Florence, R. Leshner, K.. Patel,,P. Clemens and the CINRG Investigators* Muscle Nerve 24: 787–793, 2001



MMT: reliable?

Pompe Disease

5 centers, adults

Measure	ICC	95% CI Lower-upper	Mean	MIN	MAX	SD
TMS	.929	.884 - .957	265	182	332	27.03

Myotonic Dystrophy

Quantitative Motor Assessment in Myotonic Dystrophy

J. Mathieu, H. Boivin, C.L. Richards

Can. J. Neurol. Sci. 2003; 30: 129-136

1 institution, 50 patients, adults, stratified by CTG repeats and age.

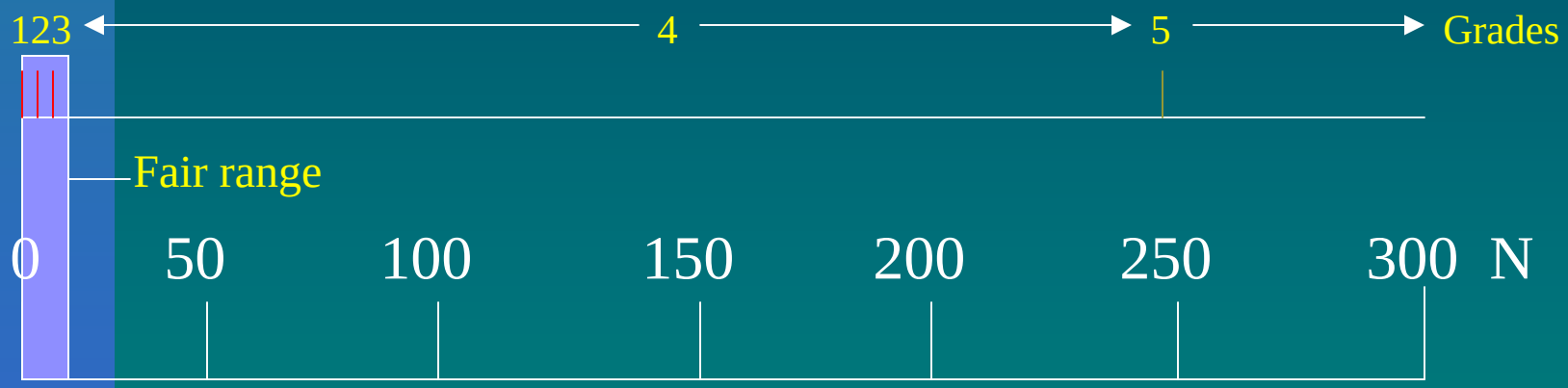
MMT of 11 muscle groups, 10 bilateraly

"The intrarater and interrater reliabilities of the total MMT score were found to be quite high (ICC=0.93 and 0.87)…"

MMT for individual muscle groups: Sensitive?

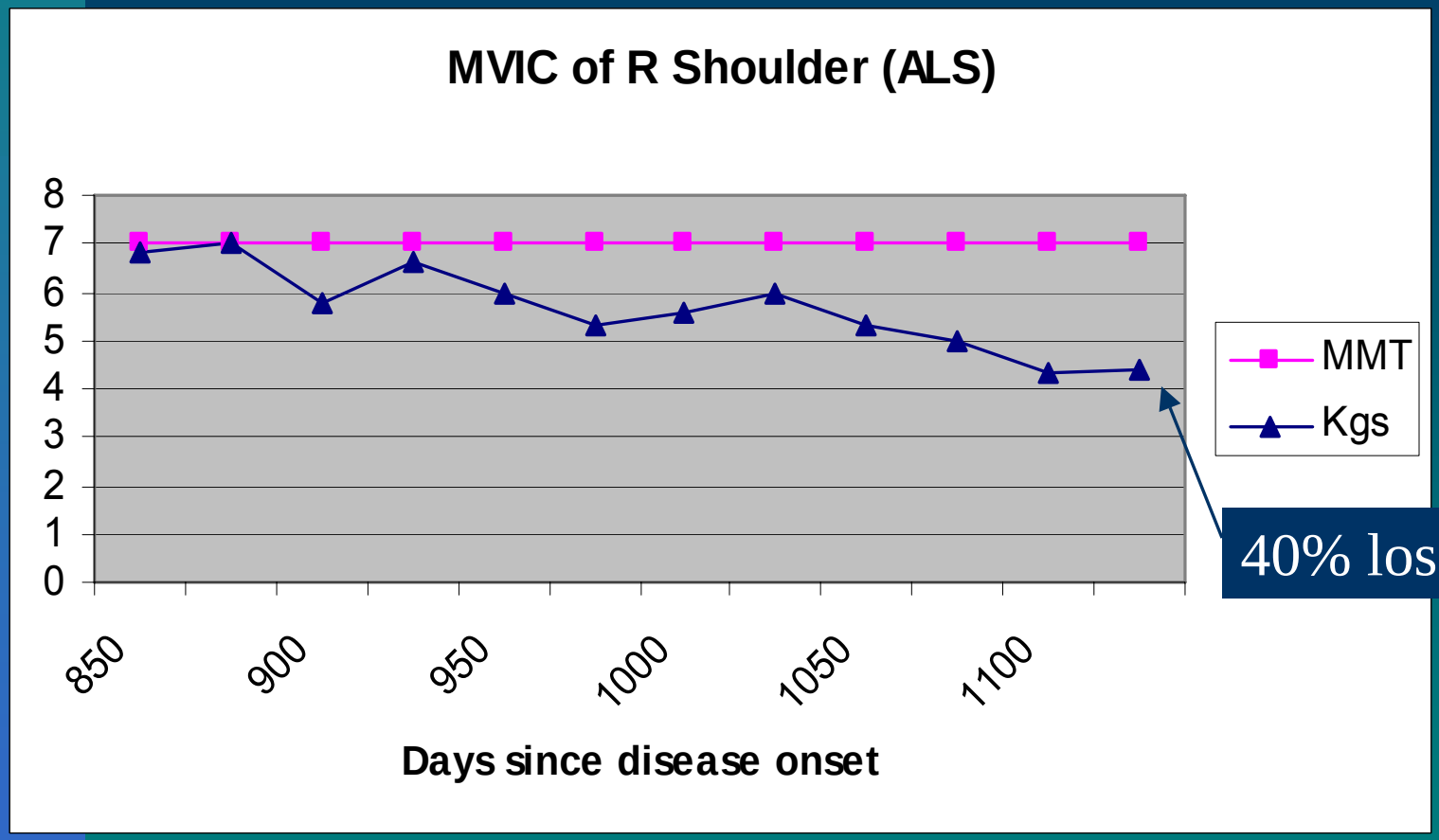
Comparison of MRC scale and Dynamometer scale

Biceps brachii uses ~3% of max Force to move through full range against gravity (F grade or 3). Grades G (4) and N (5) account for 97% of the muscle's capacity.



van der Ploeg, 1984

MMT vs. MVIC with Strain Gauge



Andres, 1989

Pompe Disease Examples

- Patient 1
  - Bilateral Knee Ext MMT Grade = 5
    - %Pred R Knee Ext = 51%
    - % Pred L Knee Ext = 48%
  - Bilateral Elbow Flx MMT Grade = 5
    - % Pred R Elbow Flx = 71%
    - % Pred L Elbow Flx = 73%
- Patient 2
  - Bilateral Knee Ext MMT Grade = 4+
    - %Pred R Knee Ext = 29%
    - % Pred L Knee Ext = 36%
  - Bilateral Elbow Flx MMT Grade = 4+
    - % Pred R Elbow Flx = 41%
    - % Pred L Elbow Flx = 71%

MMT score: Sensitive?

- Does it change with natural history? (MMT score, 34 muscle groups)
  - DMD: MMT score declined linearly by 0.4 units per year in 114 patients Brooke, Muscle Nerve 1983
  - Pompe: no change over 12 months
  - FSH: yes

Table 3 *Longitudinal changes in composite muscle strength scores*

	Composite MVICT score				Composite MMT score			
Variable	*n*	Mean	SD	*p*-value	*n*	Mean	SD	*p*-value
Change from baseline to								
6 months	48	−0.02[a]	1.06	0.91	48	−0.05	0.19	0.09
12 months	50	−0.29	0.96	0.04	47	−0.07	0.23	0.05
18 months	22	−0.44	0.89	0.03	17	−0.09	0.16	0.04
24 months	23	−0.62	1.11	0.01	20	−0.16	0.27	0.01
30 months	13	−1.32	1.81	0.02	7	−0.27	0.23	0.02
36 months	9	−1.20	0.85	0.003	8	−0.31	0.23	0.007

[a] Negative values indicate a decline in strength from baseline.

Kissel et al.

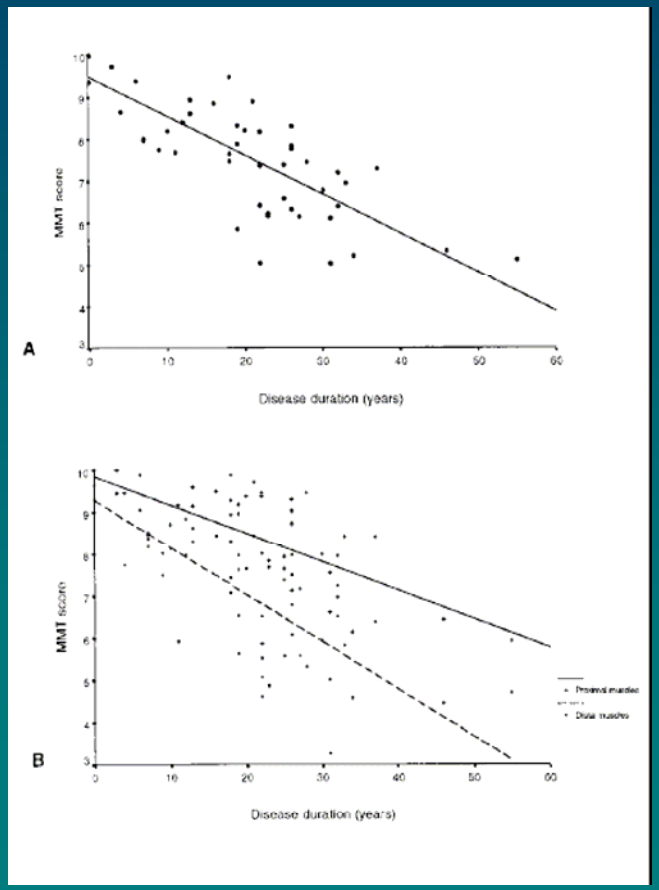
MMT score: Sensitive?

- Does it change with natural history? Myotonic Dystrophy (MMT score, 23 muscle groups; cross-sectional study) *J. Mathieu,* et al, 2003

Muscle group	Gender	Score			Slope of linear regression			Decline per year
		Mean	±SD	r^2	β	(95% CI)	p	of disease duration (%)
A) Manual muscle testing (MMT)								
Proximal muscles	Men	8.57	1.30	0.40	-0.08	(-0.12; -0.04)	0.0004	0.80
	Women	8.50	1.09	0.66	-0.06	(-0.08; -0.04)	<0.0001	0.61
	Total	8.54	1.20	0.49	-0.07	(-0.09; -0.05)	<0.0001	0.71
Distal muscles	Men	7.07	1.94	0.67	-0.15	(-0.19; -0.11)	<0.0001	1.54
	Women	7.16	1.60	0.65	-0.09	(-0.12; -0.06)	<0.0001	0.99
	Total	7.11	1.78	0.61	-0.11	(-0.14; -0.09)	<0.0001	1.19
All muscles	Men	7.69	1.56	0.62	-0.12	(-0.15; -0.08)	<0.0001	1.23
	Women	7.67	1.32	0.74	-0.08	(-0.10; -0.06)	<0.0001	0.86
	Total	7.68	1.44	0.64	-0.09	(-0.11; -0.07)	<0.0001	0.95

MMT score: Sensitive?

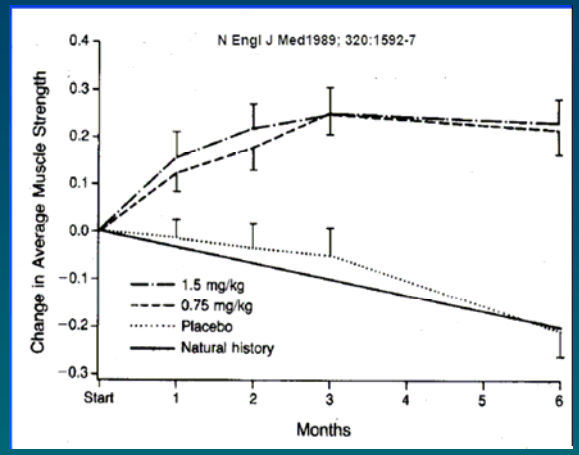
- MMT shows linear deterioration

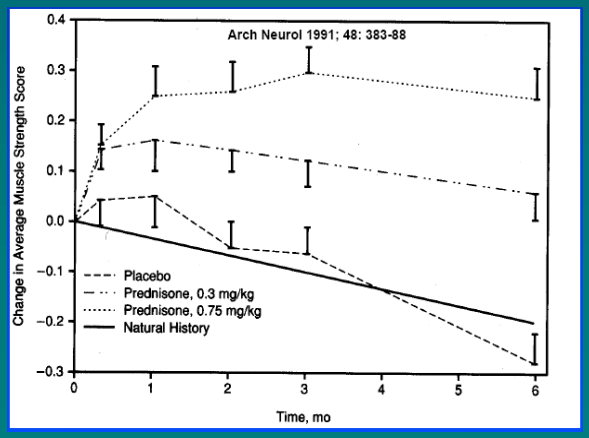


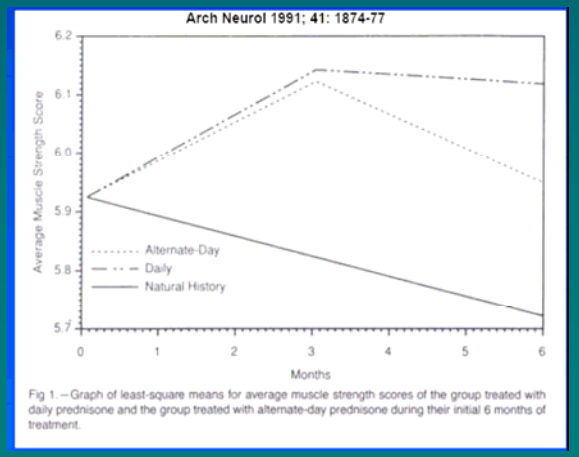
- MMT score correlats with CTG repeats
  - " The muscle strength as expressed by the MMT total score is highly correlated with the disease duration (Pearson r = -0.79, p<0.0001). As the disease duration is also highly correlated with the CTG-repeat sizes (Spearman r= -.56, p<0.0001), it is not surprising to find a significant correlation between the MMT total score and the CTG-repeat sizes (Spearman r = -0.56, p<0.0001)."

MMT score sensitivity

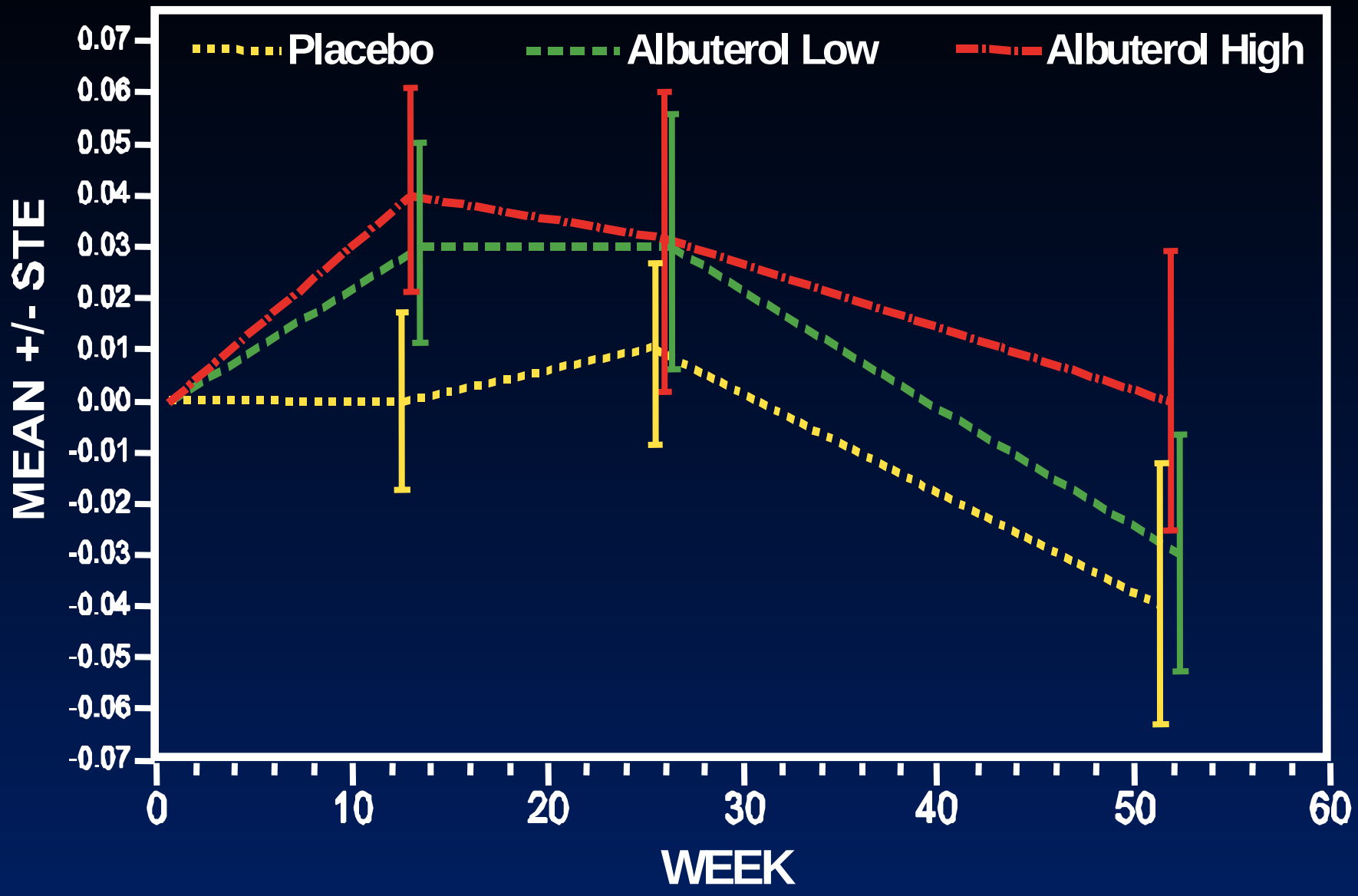
- Does it change with interventions?
  - Only intervention to date to show an effect in strength are corticosteroids in DMD







Change in Avg MMT



MMT score: sensitivity

- But, more recent experience shows change in natural history data (younger patients?) and shows MMT insensitive to smaller effects
  - CINRG randomized controlled trial of creatine and glutamine in Duchenne muscular dystrophy. Escolar DM, Buyse G, Henricson E, Leshner R, Florence J, Mayhew J, et al. Ann Neurol. 2005 Jul;58(1):151-5.
  - A randomized efficacy and safety trial of oxandrolone in the treatment of Duchenne dystrophy G.M. Fenichel, MD; R.C. Griggs MD; J. Kissel, MD et al. NEUROLOGY 2001;56:1075–1079

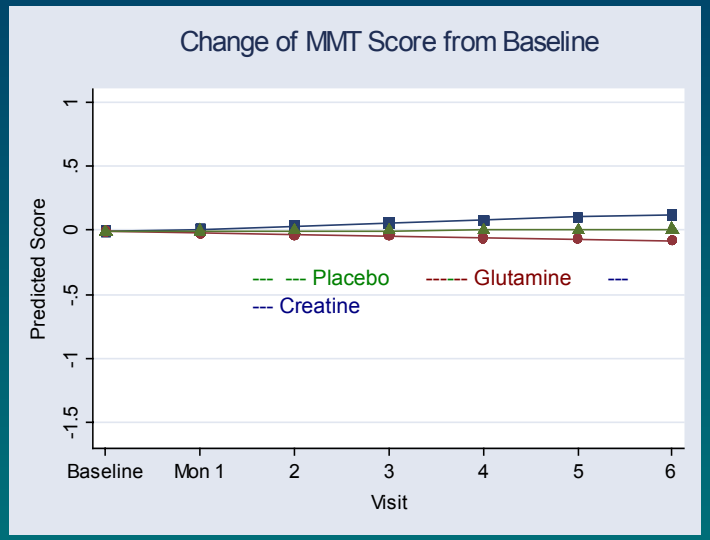


Table 2 Mean change in manual muscle strength test scores

	Treatment group							
	Oxandrolone				Placebo			
Week	N	Mean (SE)	Min	Max	N	Mean (SE)	Min	Max
4	24	0.015 (0.061)	−0.750	0.632	24	−0.063 (0.059)	−0.603	0.632
12	26	0.067 (0.063)	−0.5000	0.632	25	−0.182 (0.058)	−0.656	0.221
25	26	0.031 (0.085)	−0.838	0.750	24	−0.135 (0.091)	−0.946	0.879
26	26	0.039 (0.073)	−0.618	0.838	24	−0.145 (0.080)	−1.096	0.606
Average, weeks 25 and 26	26	0.035 (0.077)	−0.632	0.735	24	−0.140 (0.082)	0.982	0.742

MMT score: Clinical meaningfulness

- It is unknown what magnitude of change in MMT score correspond to a change in function, QOL, or clinically significant effect
- However: A 0.4 points change with prednisone at 6 months is “noticeable” by patient, parents and doctors
- Should any new therapy shoot for this change or higher?
- What does a 0.1 MMT score means in QMT score?

MMT score: Feasibility

- Standardized method of testing
- Time and cost effective
- No equipment needed
- Needs significant training
- Moderate reliability in multi-institution trial setting (>4)

Isometric muscle testing Manual muscle testing

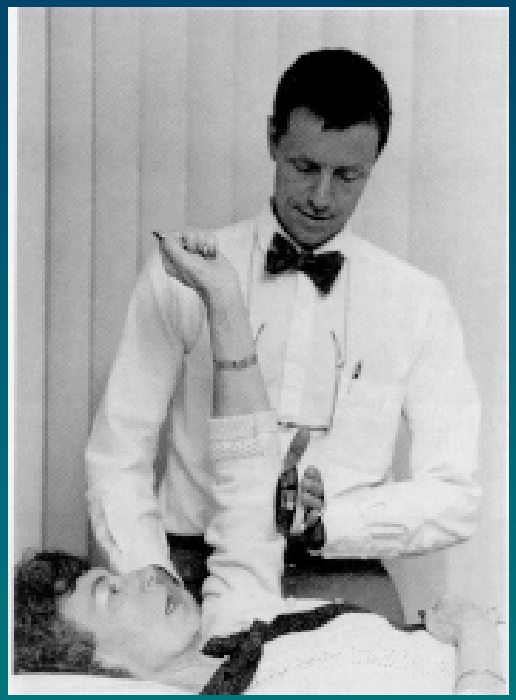
- Pros
  - Time and cost effective
  - Reliable as a total muscle score and with small number of centers/CEs
  - Sensitive showing progression in DMD , FSH and MD through natural history studies
  - Shows large effects in steroid trials in DMD
- Cons
  - Poor to moderate reliability with individual muscle groups
  - Less reliable and need more training for large multi-center setting
  - Low sensitivity in showing change especially in stronger muscle groups (MRC grade 4- and up)
  - Ordinal scale: less precision
- Missing
  - Consensus on what change in score is clinically meaningful?

Isometric muscle testing
2- Hand-Held Dynamometers

- Principles:
  - Device is interposed between the hand and the limb to be tested
  - Patient performs a “break test”
- Device is usually:
  - Oil-filled chamber transmitting to a pressure gauge, or
  - Strain gauge

Hand-Held Dynamometers

Often performed in standard MMT positions



Hand-Held Dynamometers Reliable?

- Stuberg, 1988
  - 28 male children
    - 14 DMD
    - 14 Age matched normals
    - Range (6-13 years)
- Test-retest correlation
  - Normals: correlations ranged from .96 to .99
    - Not all tests could be performed as they were too strong for device
  - DMD: correlations ranged from .83 to .99

Hand-Held Dynamometers

- **Sensitive?**
  - Has not been used in natural history studies or clinical trials of muscle disorders
  - Not sensitive to change in stronger muscles
  - Normative values exist for children
    - Backman E Methods for measurement of muscle function. Methodological aspects, reference values for children, and clinical applications. Scand J Rehabil Med Suppl. 1988;20:9-95.
    - Connolly (unpublished) for DMD
- **Clinical meaningfulness? Validity?**
- **Feasibility**
  - Inexpensive
  - Training probably similar than that for MMT

Hand-Held Dynamometers

- Limitations for clinical trials:
  - Strong muscle groups can't be tested
  - Stabilization can be difficult
  - Blinding measurements is an extra step
  - Data needs to be manually entry

Summary of Hand-Held dynamometer

- Benefits
  - Quantitative
  - Easy to use
  - cheap
  - Portable
- Drawbacks
  - Less reliable
  - Less sensitive in stronger muscles
  - Manual data entry

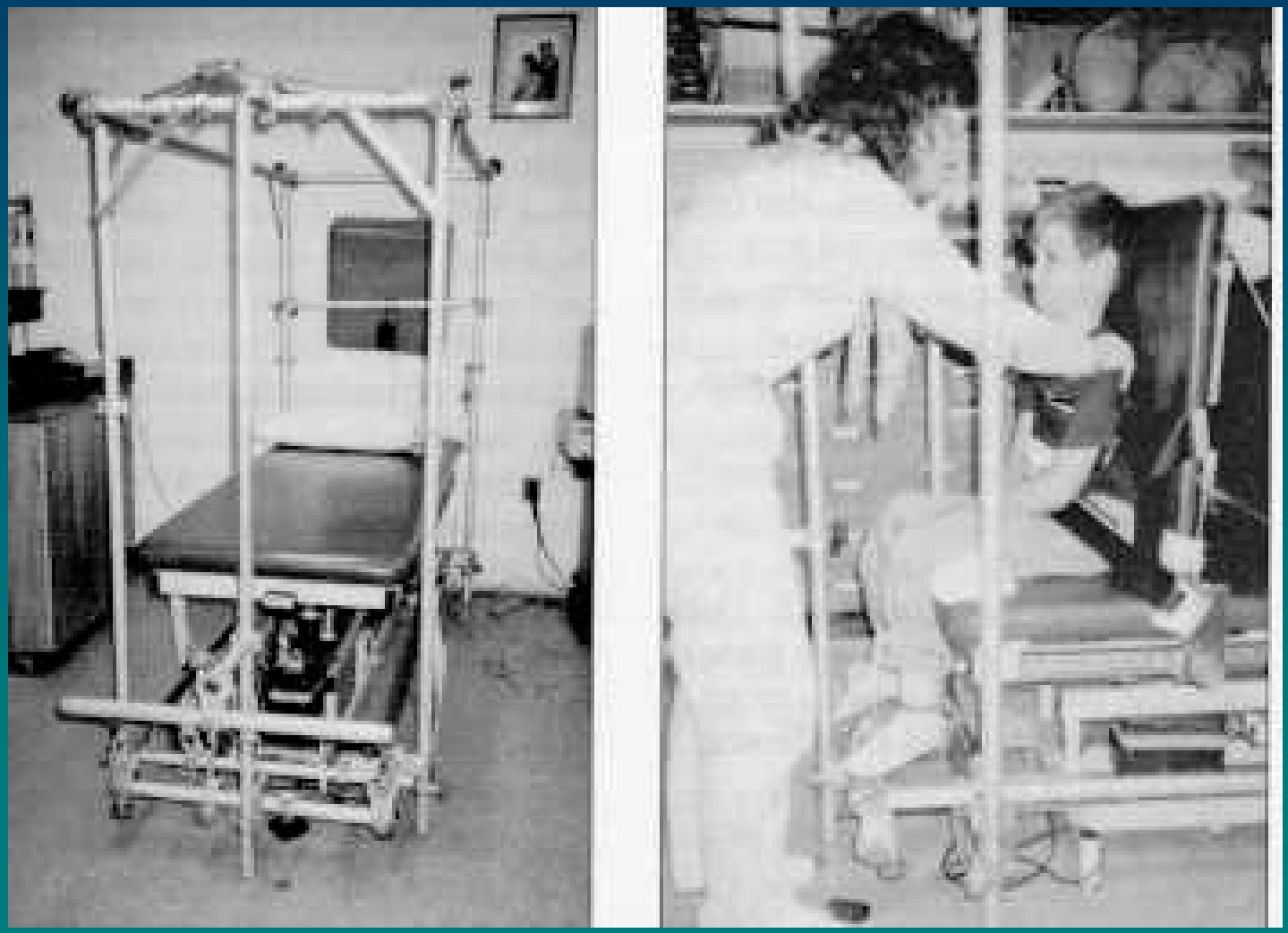
Isometric muscle testing
3- Strain Gauge Dynamometers

- Load causes a deformation of material
- Strain gauge measures the deformation
- Change in voltage can be calibrated to represent change in force
- Frequently used on bars or rings in hand-held dynamometers

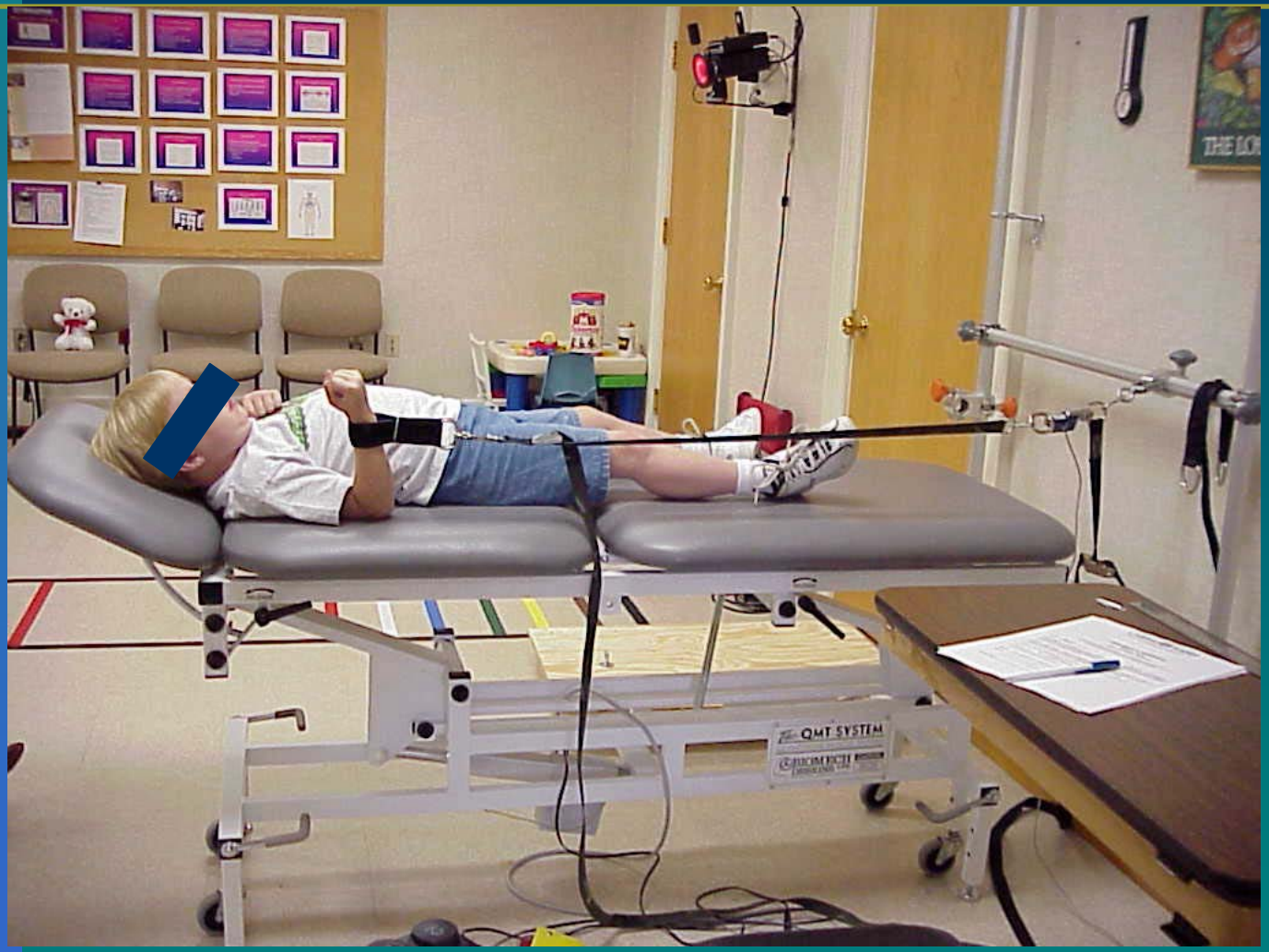
Strain Gauge Methods

- Andres, 1986
  - Development of the Tufts Quantitative Neuromuscular Exam (TQNE)
  - One component – isometric test (MVIC)
  - Equipment:
    - Table
    - Attached bars
    - Cables, straps, load cell
    - A/D board – computer

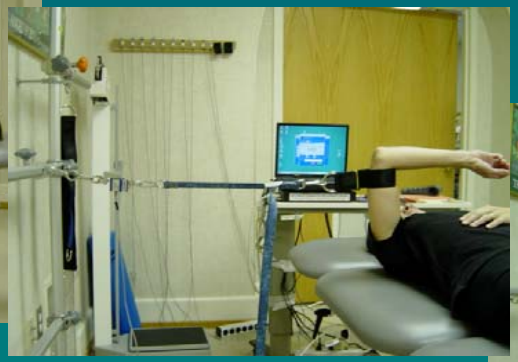
TQNE



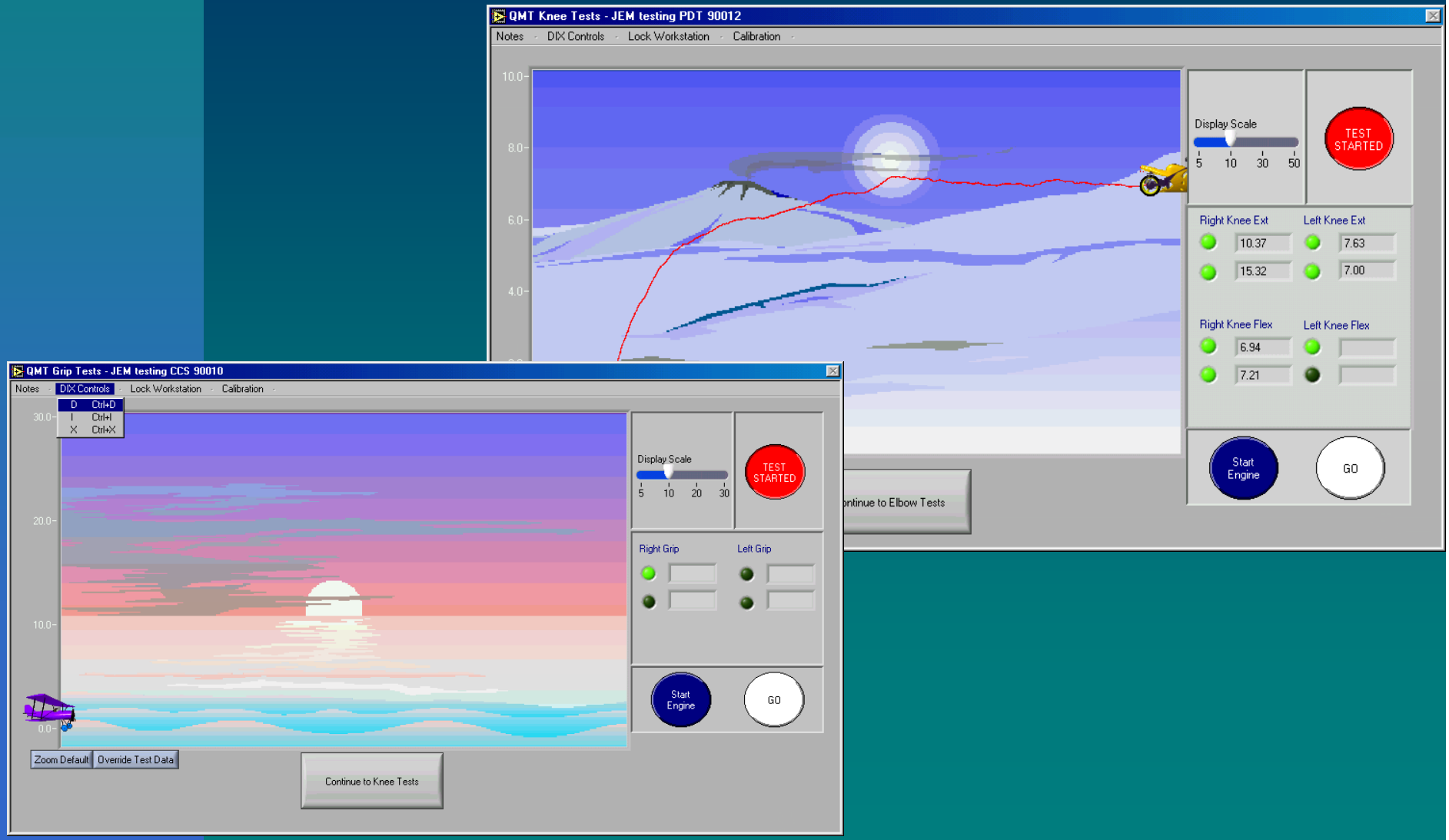
CINRG Quantitative Muscle System



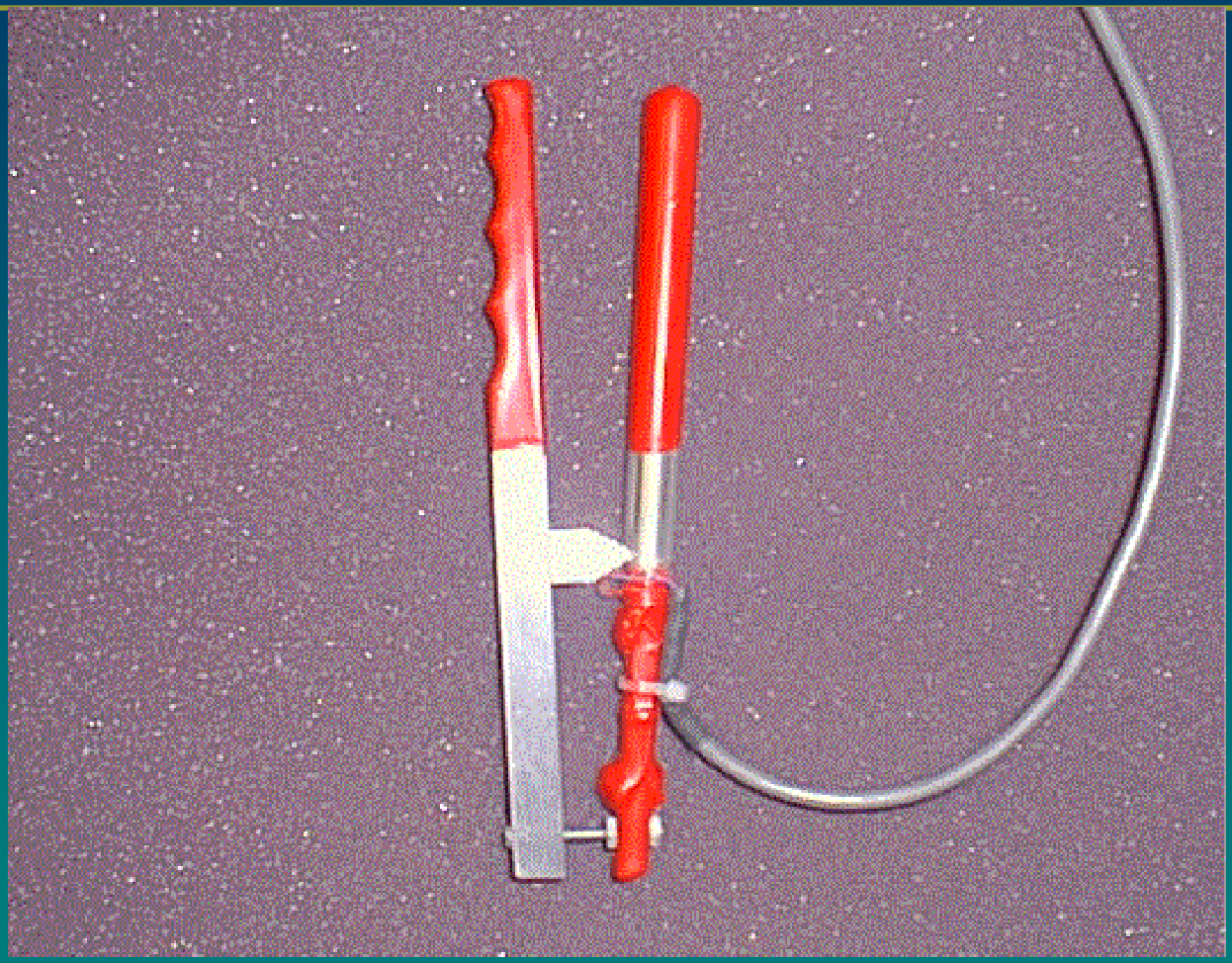
CQMS



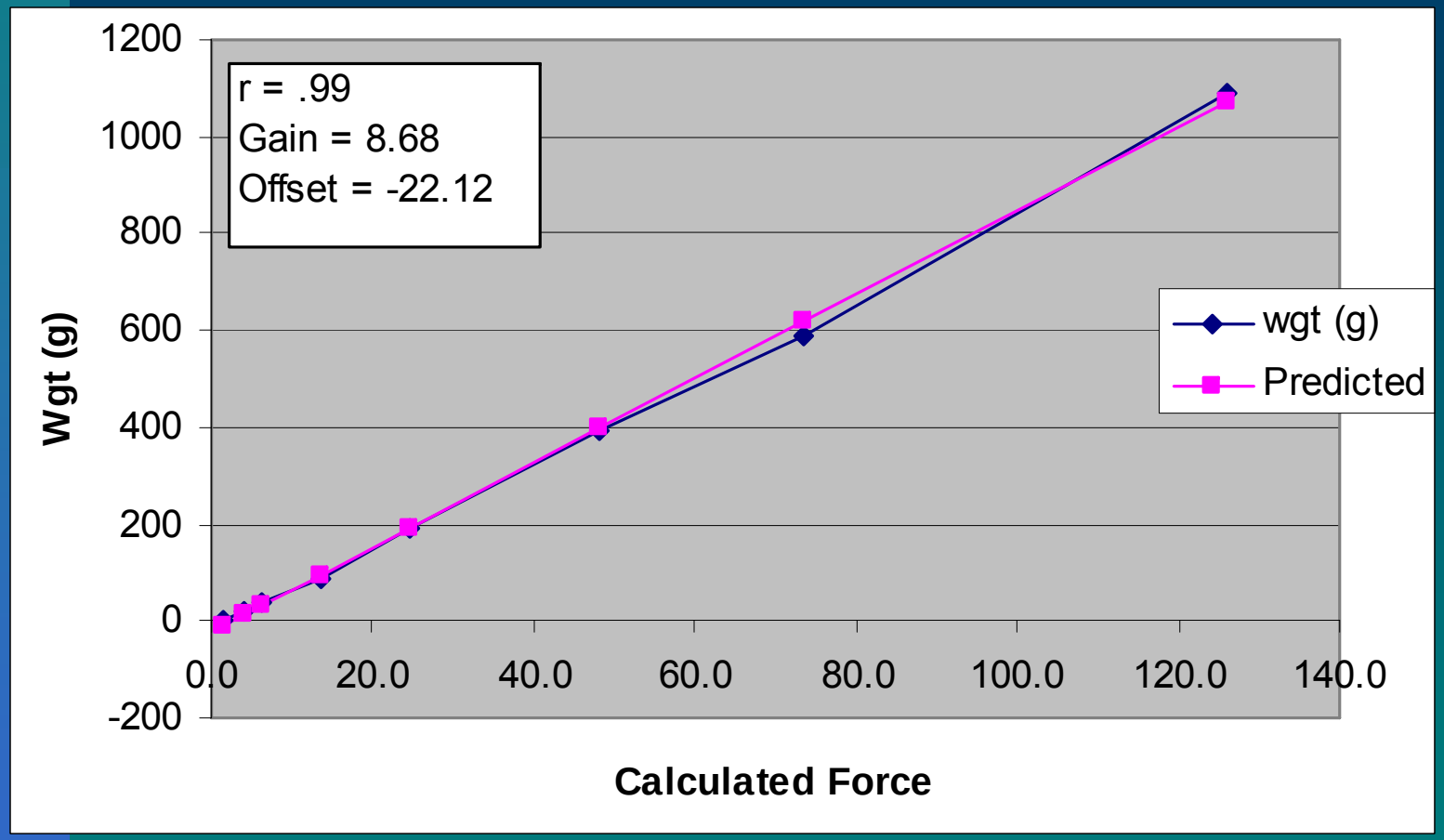
CQMS Data Collection Screen – Knee



Grip Dynamometer

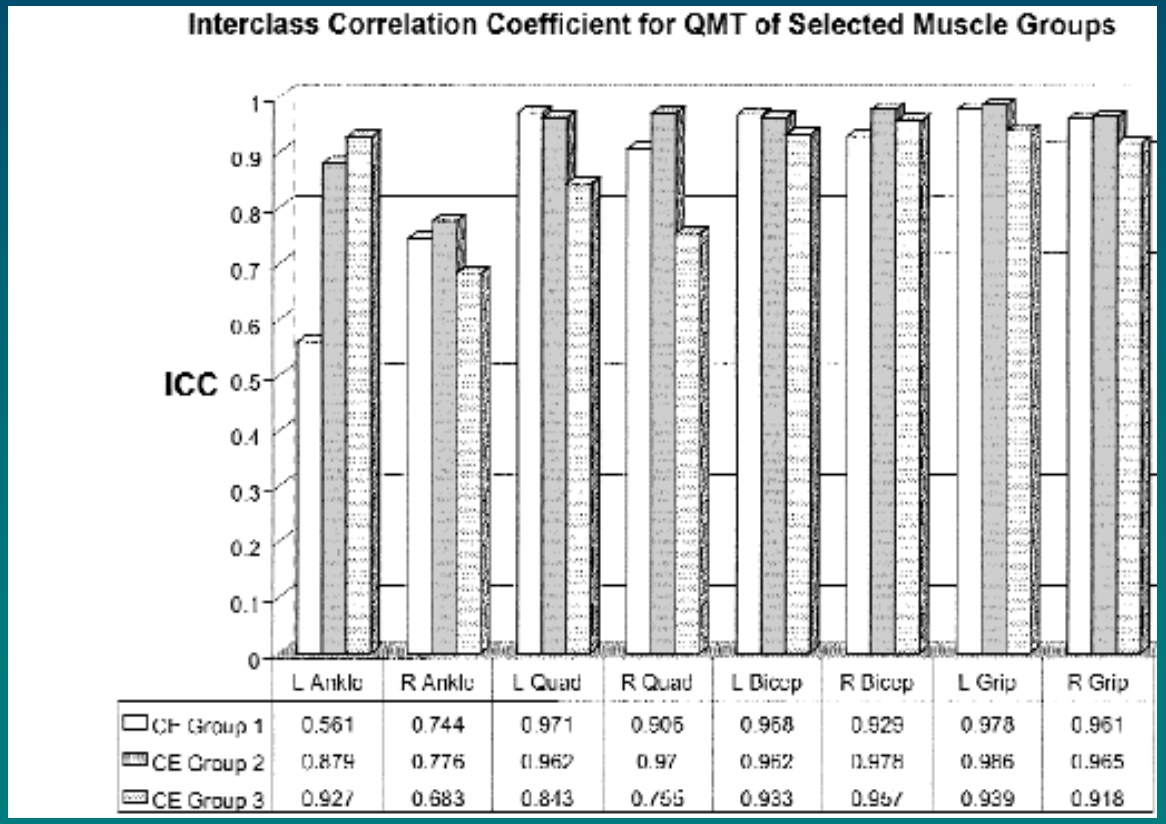


Actual vs. Predicted Force



QMT/CQMS/ MVIC Reliable?

CINRG Muscle Nerve 24: 787–793, 2001



	L Ankle	R Ankle	L Quad	R Quad	L Bicep	R Bicep	L Grip	R Grip
CE Group 1	0.561	0.744	0.971	0.906	0.968	0.929	0.978	0.961
CE Group 2	0.879	0.776	0.962	0.97	0.952	0.978	0.986	0.965
CE Group 3	0.927	0.683	0.843	0.755	0.933	0.967	0.939	0.918

CQMS reliability in 32 DMD 4-16 years old

	Min - Max	1 vs 2 Gold Standard	95% CI	SEM	Min Det Change 68% CI	Min Det Change 90% CI
TMS	140 – 302 points	.83	.32 - .94	11.54	16 points	26 points
Total QMT	33.68 – 278.55 lbs	.97	.93 - .98	8.49	12 lbs	19.8 lbs
R Grip	2.60 – 35.59 lbs	.87	.74 - .94	2.31	3.27 lbs	5.4 lbs
L Grip	2.52 – 37.00 lbs	.95	.89 - .97	1.41	1.99 lbs	3.28 lbs
R Knee Ext	1.19 – 52.30 lbs	.90	.79 - .95	3.00	4.24 lbs	7.0 lbs
L Knee Ext	1.34 – 56.69 lbs	.91	.82 - .96	2.96	4.19 lbs	6.91 lbs
R Knee Flex	2.35 – 28.43 lbs	.89	.79 - .95	1.70	2.40 lbs	3.96 lbs
L Knee Flex	2.01 – 26.23 lbs	.79	.55 - .90	2.31	3.27 lbs	5.4 lbs
R Elbow Ext	.91 – 16.42 lbs	.92	.76 - .97	.89	1.26 lbs	2.08 lbs
L Elbow Ext	1.06 – 15.60 lbs	.88	.70 - .95	1.21	1.71 lbs	2.82 lbs
R Elbow Flex	1.14 – 18.48 lbs	.95	.90 - .98	.79	1.11 lbs	1.83 lbs
L Elbow Flex	.86 – 17.52 lbs	.97	.93 - .98	.61	.86 lbs	1.42 lbs

CQMS reliability in Late–onset Pompe Disease

Measure	ICC	95% CI Lower-upper	Mean	MIN	MAX	SD
R grip	.973	.955 - .984	74.68	22.44	136.53	23.38
L grip	.974	.957 - .984	72.05	20.74	121.49	24.72
R elbow flx	.987	.978 - .992	33.66	5.62	68.70	15.72
L elbow flx	.991	.985 - .995	35.88	7.27	70.02	16.08
R elbow ext	.938	.899 - .963	23.44	3.55	62.53	10.73
L elbow ext	.946	.911 - .968	25.54	3.99	57.71	10.70
R shldr add	.939	.899 - .963	12.27	0.00	38.56	8.64
L shldr add	.929	.883 - .957	14.26	1.07	36.04	8.39
R knee flx	.980	.966 - .988	21.38	2.18	66.31	15.06
L knee flx	.981	.969 - .989	21.96	1.53	62.58	15.03
R knee ext	.969	.948 - .981	35.59	3.10	98.11	23.35
L knee ext	.978	.964 - .987	37.30	2.84	105.64	24.18
R hip add	.957	.929 - .974	18.59	1.43	56.10	12.81
L hip add	.936	.894 - .961	18.58	1.49	48.65	12.36

QMT/CQMS/ MVIC Sensitive?

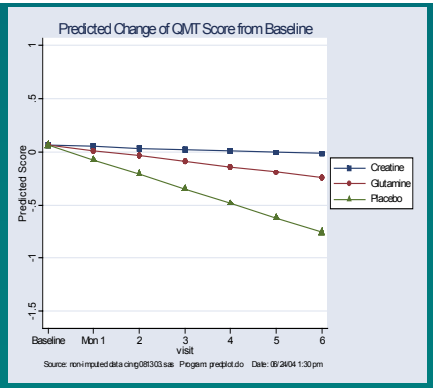
- To change over time in natural history?
  - FSH study group

Table 3 *Longitudinal changes in composite muscle strength scores*

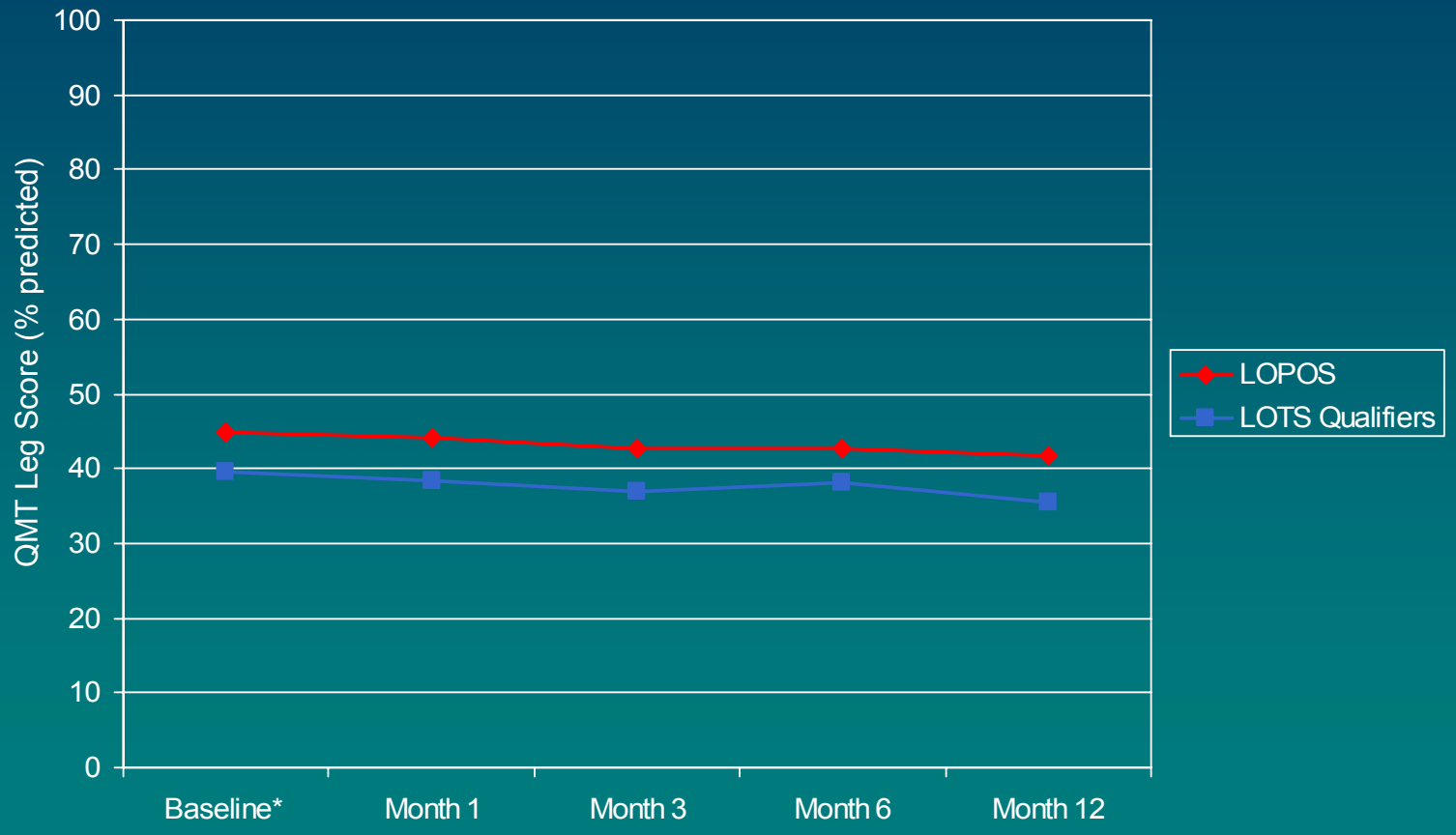
	Composite MVICT score				Composite MMT score			
Variable	*n*	Mean	SD	*p*-value	*n*	Mean	SD	*p*-value
Change from baseline to								
6 months	48	−0.02[a]	1.06	0.91	48	−0.05	0.19	0.09
12 months	50	−0.29	0.96	0.04	47	−0.07	0.23	0.05
18 months	22	−0.44	0.89	0.03	17	−0.09	0.16	0.04
24 months	23	−0.62	1.11	0.01	20	−0.16	0.27	0.01
30 months	13	−1.32	1.81	0.02	7	−0.27	0.23	0.02
36 months	9	−1.20	0.85	0.003	8	−0.31	0.23	0.007

[a] Negative values indicate a decline in strength from baseline.

  - DMD by CINRG



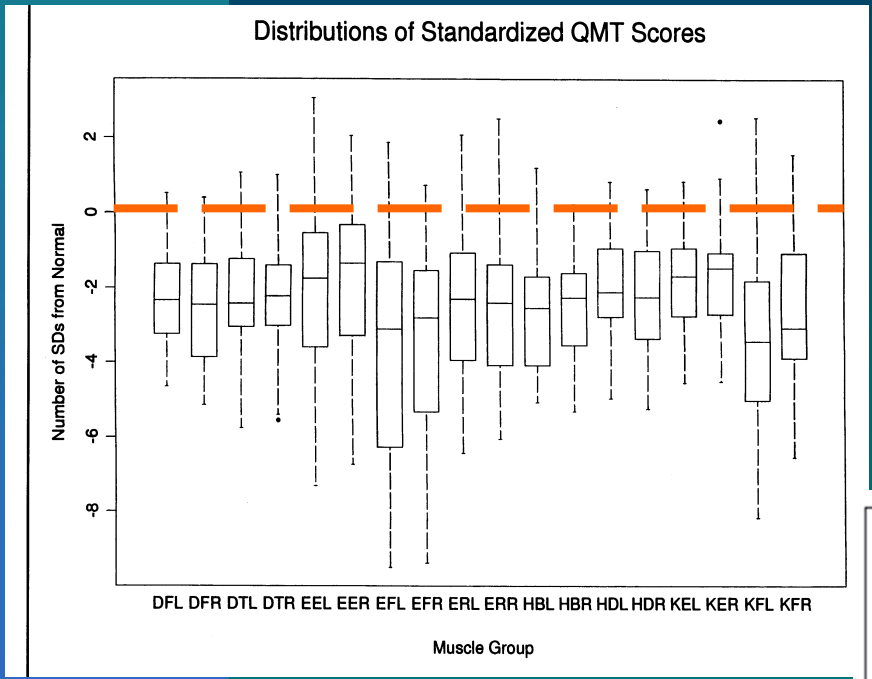
QMT Leg Score in LOPOS



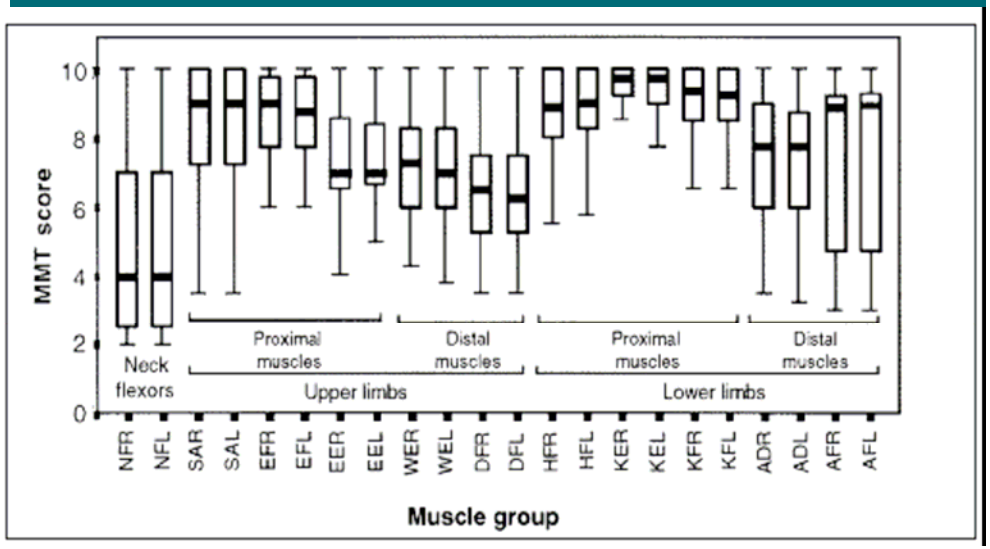
* Values from Day 1 of Baseline visit

QMT/CQMS/ MVIC
Does it describe clinical distribution of weakness?

MVICT Muscle Scores - FSHD



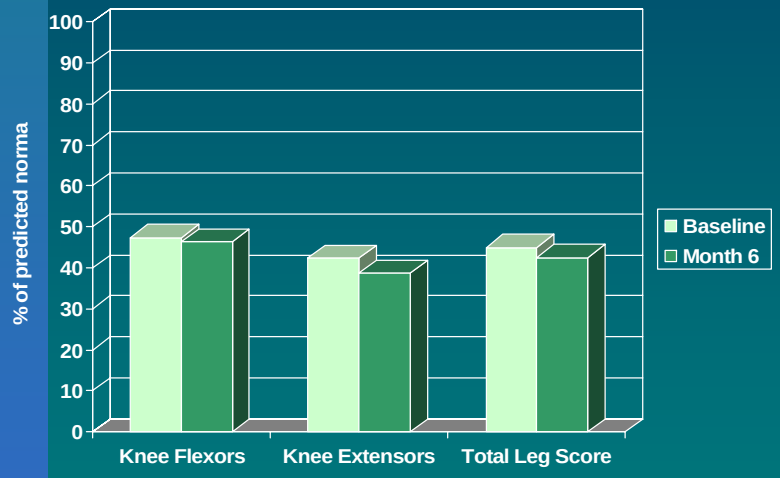
QMT Muscle Scores - MD



QMT/CQMS/ MVIC
Does it describe clinical distribution of weakness?

LOPOS study

Lower Extremity Strength by QMT



Baseline (n = 58)
7.0 – 115.0%

Baseline (n = 58)
5.2 – 89.5%

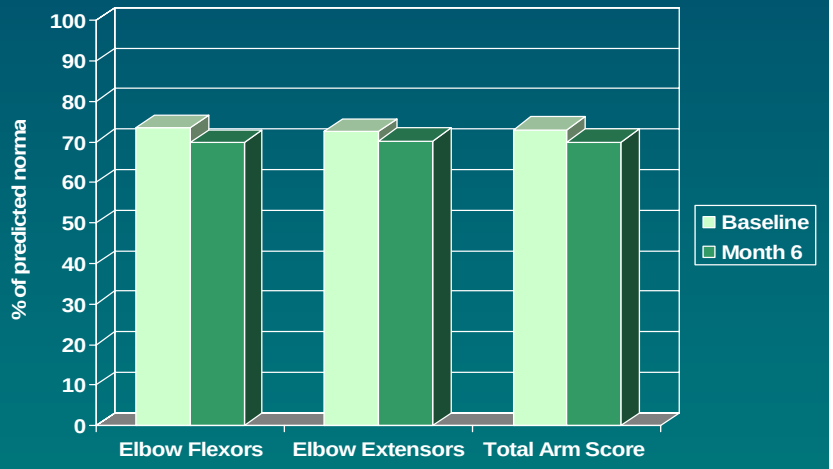
Baseline (n = 58)
9.9 – 99.5%

Month 6 (n = 58)
6.5 – 109.6%

Month 6 (n = 56)
4.2 – 98.2%

Month 6 (n = 58)
8.2 – 93.4%

Upper Extremity Strength by QMT



Baseline (n = 58)
18.4 – 127.5%

Baseline (n = 58)
20.7 – 126.0%

Baseline (n = 58)
19.6 – 123.5%

Month 6 (n = 58)
18.1 – 123.6%

Month 6 (n = 56)
24.2 – 116.5%

Month 6 (n = 58)
21.1 – 109.6%

QMT: does it respond to interventions?

Prednisone in DMD daily for 10 days of the month. 5-8 years old, ambulant

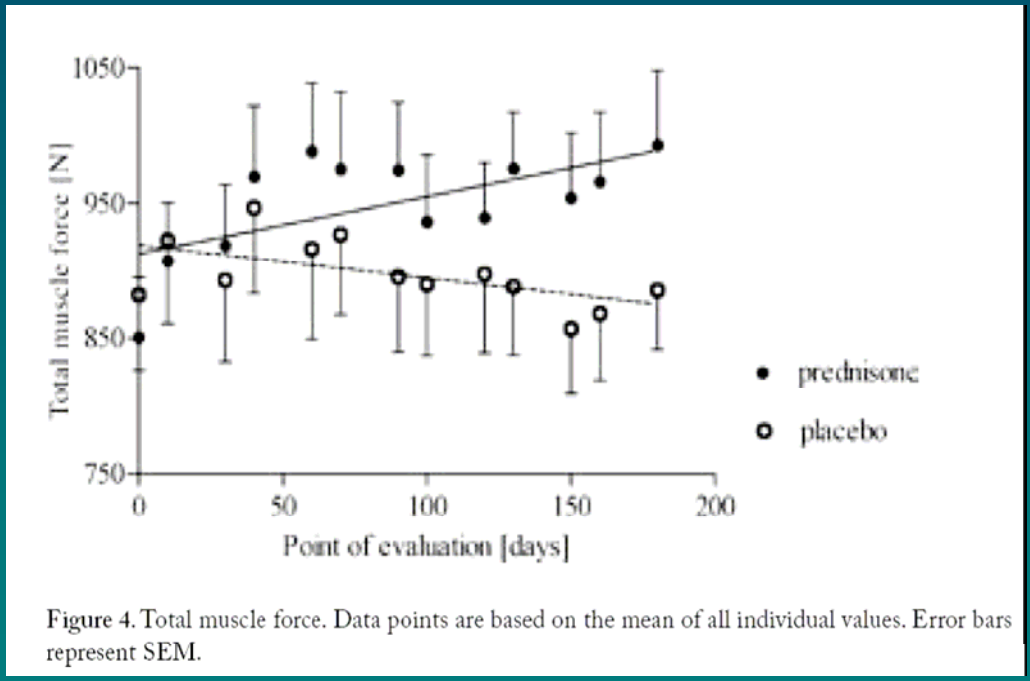
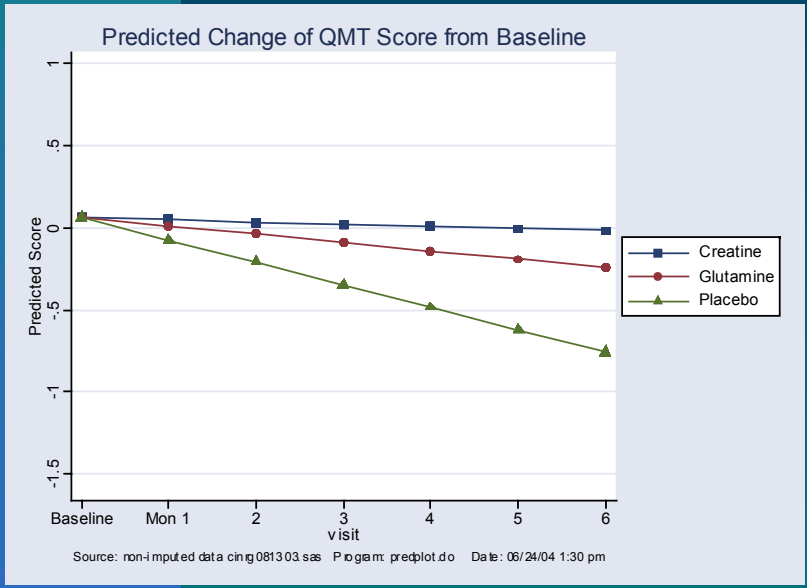


Figure 4. Total muscle force. Data points are based on the mean of all individual values. Error bars represent SEM.

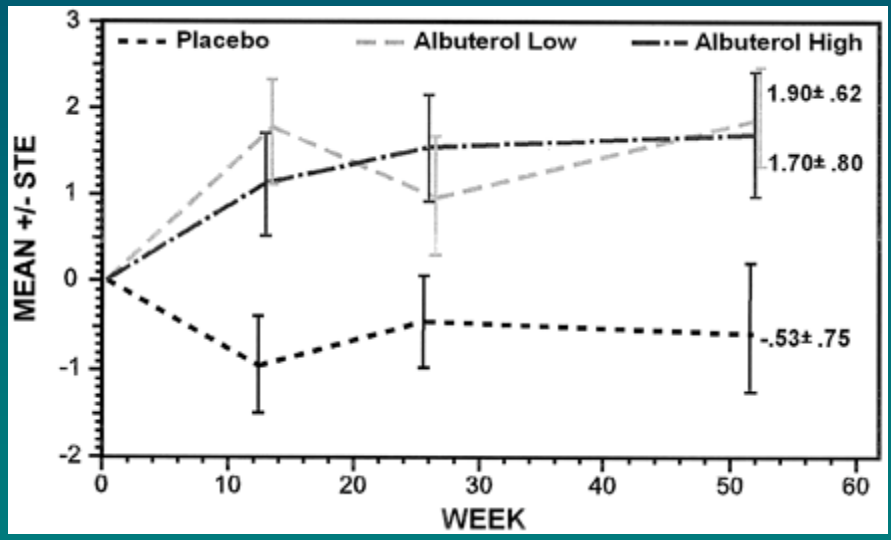
Beenakker et al. Arch Neurol. 2005 Jan;62(1):128-32

QMT: does it respond to interventions?

CINRG. Creatine and Glutamine trial

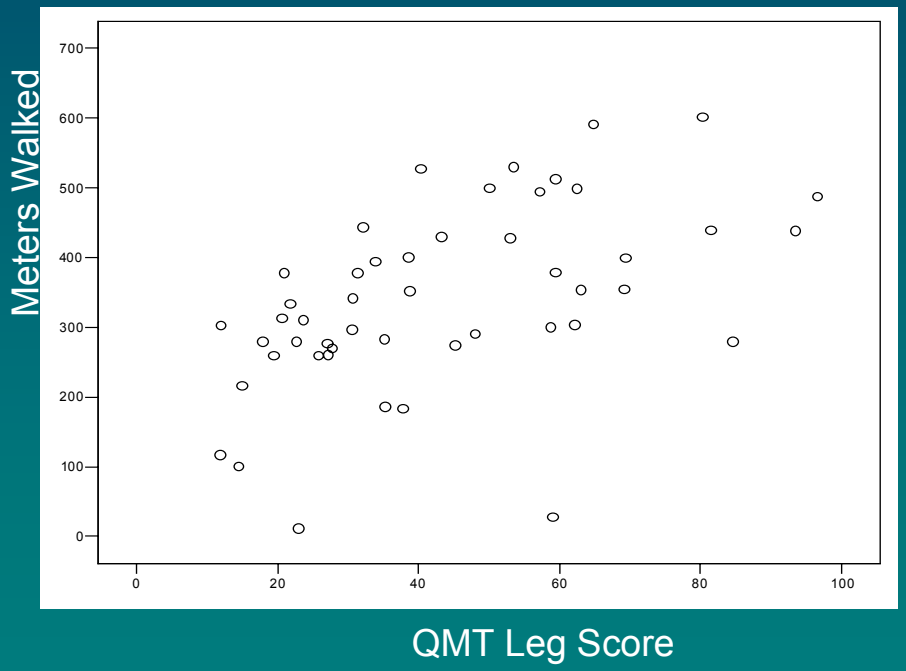


FSH group. Albuterol trial Change in grip

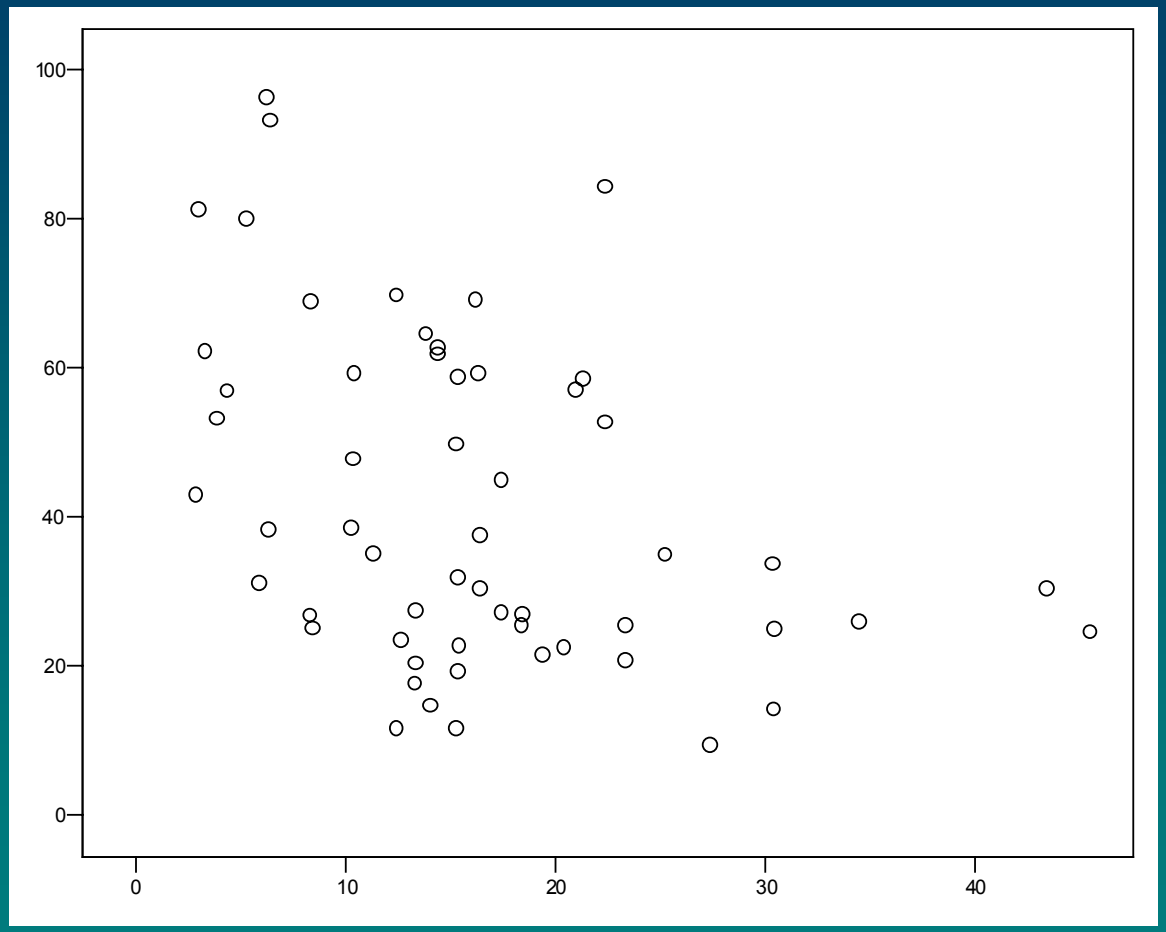


QMT validity as surrogate measure Does it correlate with function? QOL?

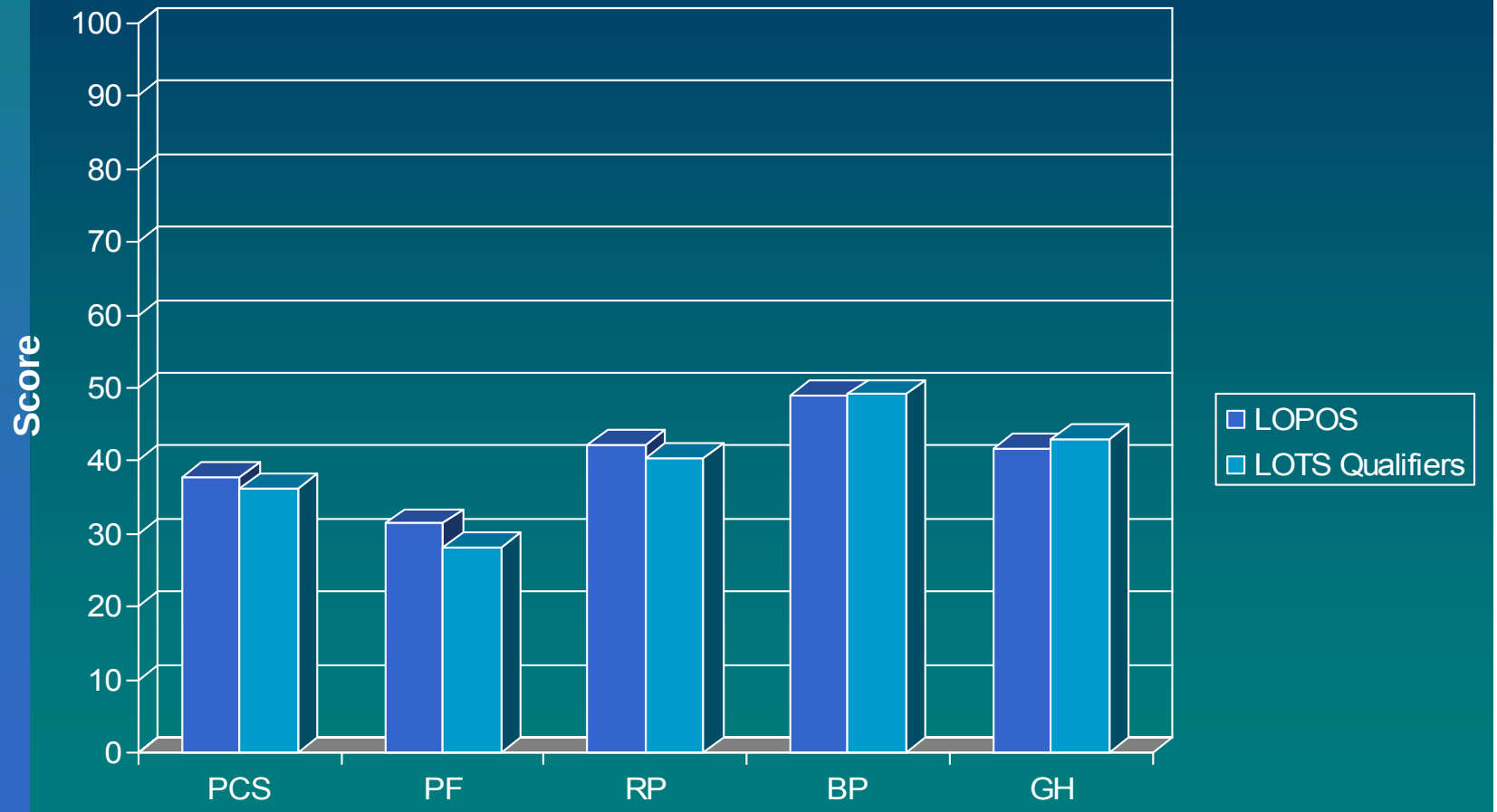
- LOPOS study: QMT Leg Score relationship to 6MWT Distance



LOPOS: Disease Severity and Lower Extremity Muscle Weakness



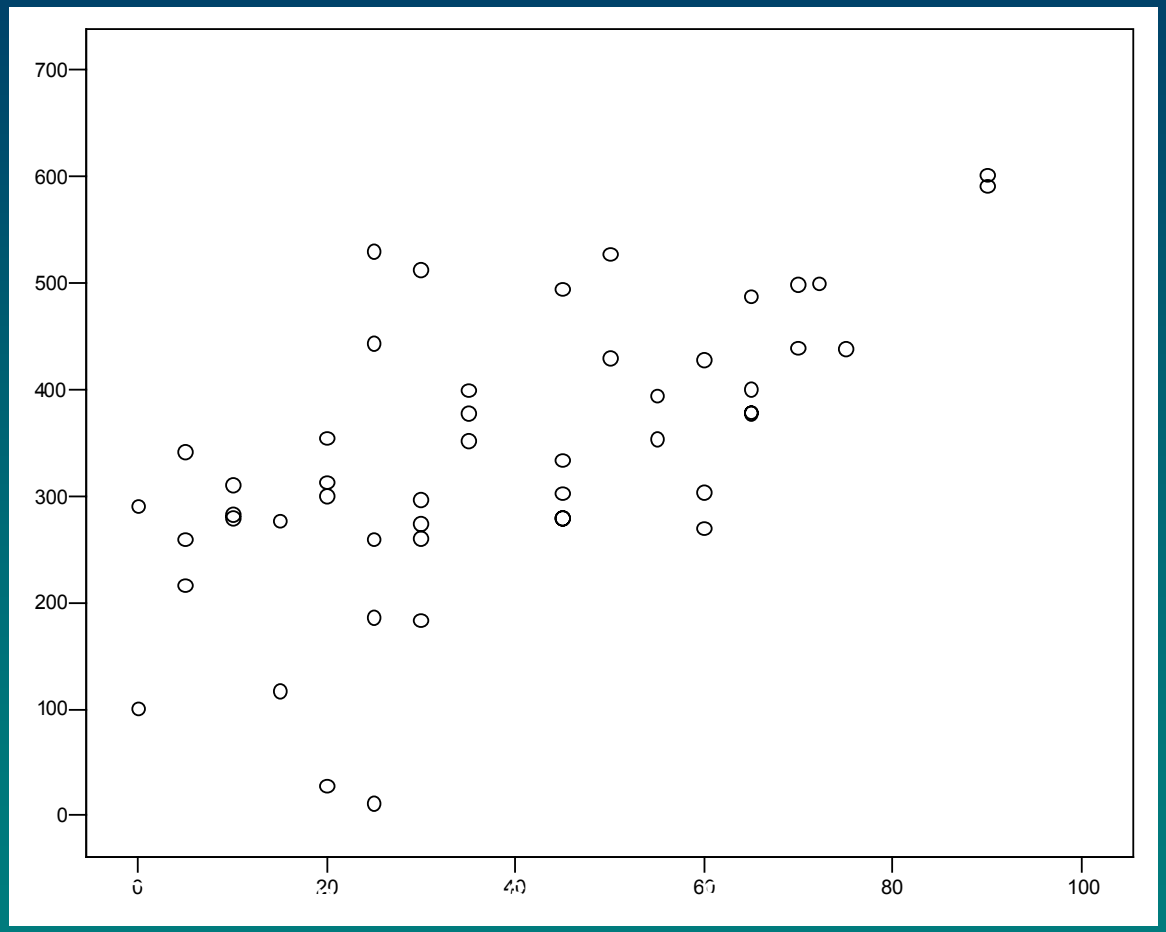
LOPOS: SF-36 PCS Score



PCS - Physical Component Summary Score **Scale**
PF - Physical Functioning
RP- Role Physical
BP – Bodily Pain
GH – General Health

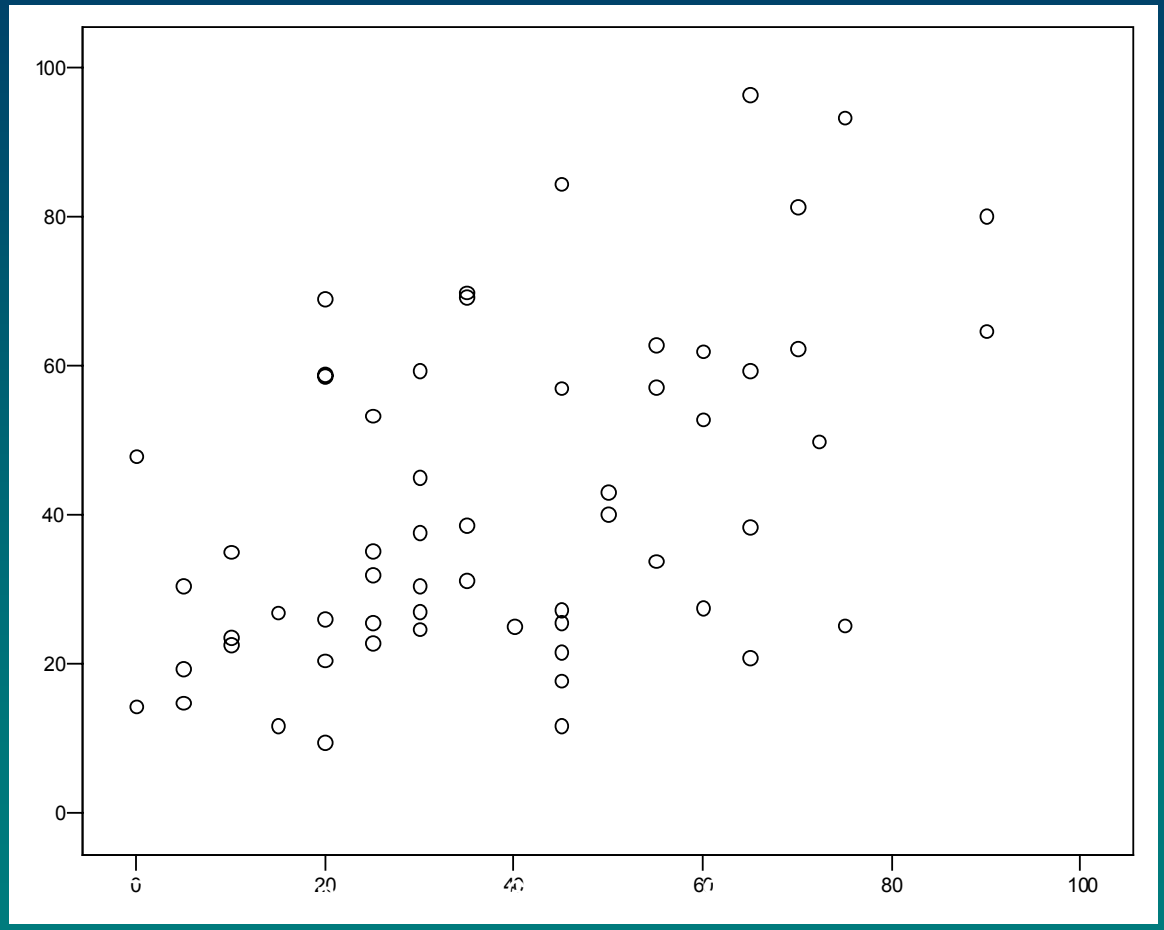
Results based on LOPOS Month 12 study visit

SF-36 PF Score: Relationship to 6MWT Distance



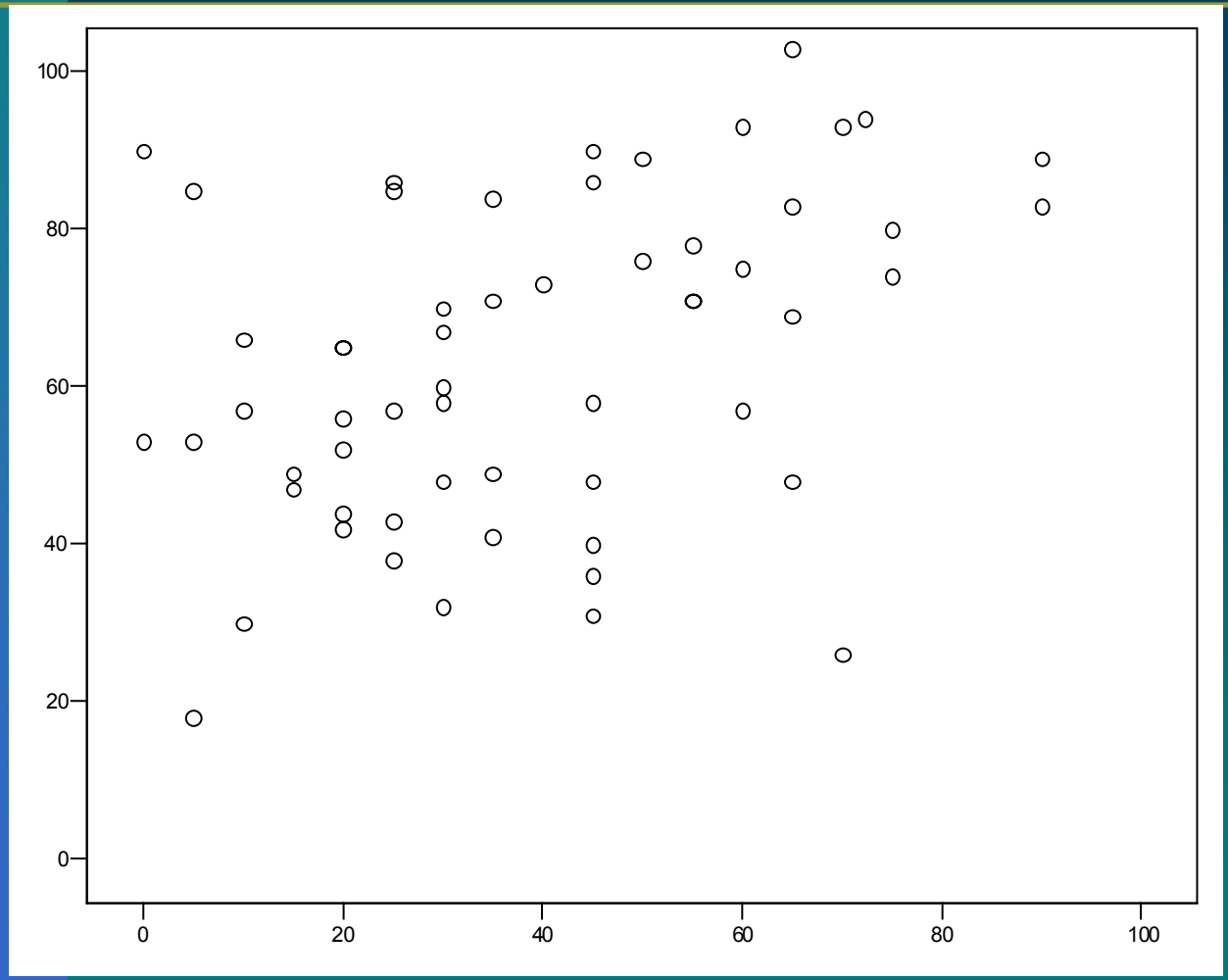
Results based on LOPOS Month 12 study visit

SF-36 PF Score: Relationship to QMT Leg Score



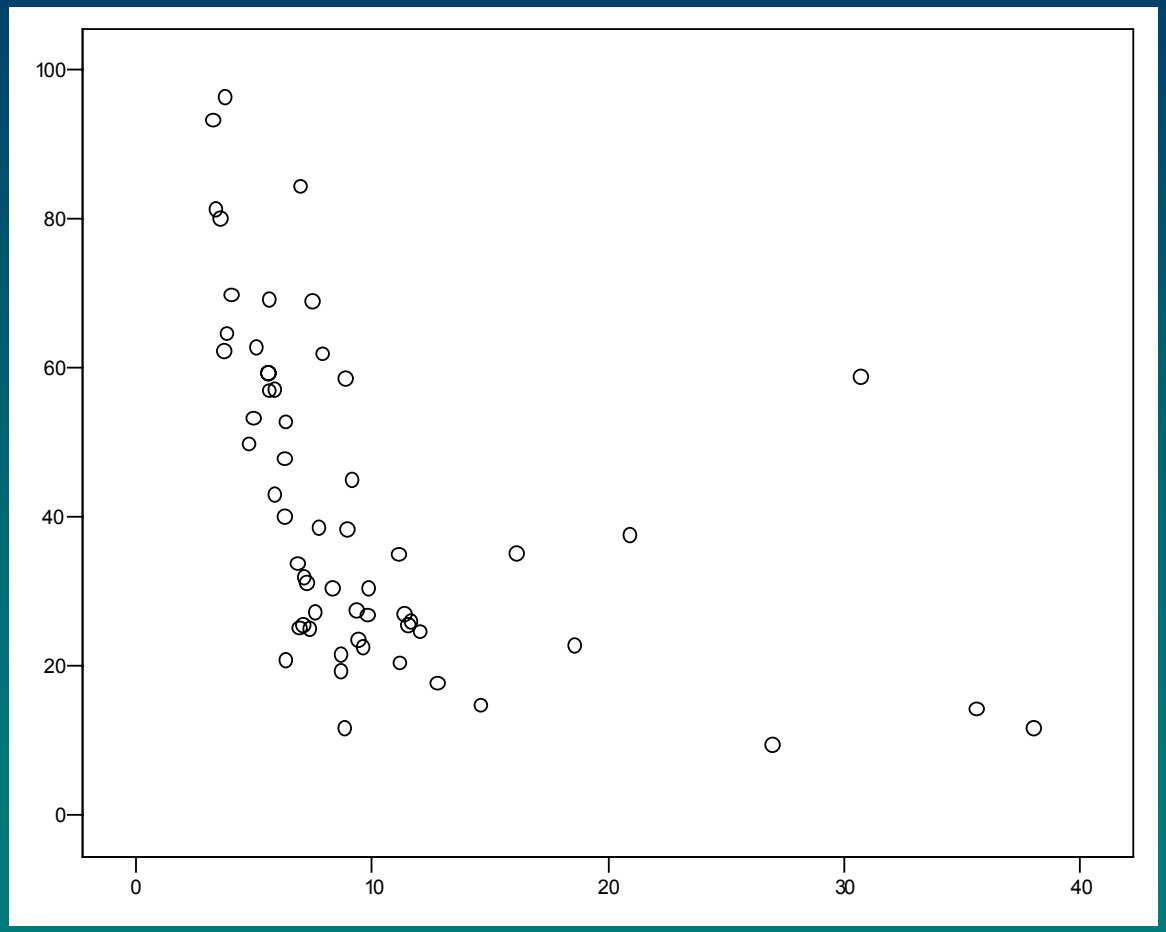
Results based on LOPOS Month 12 study visit

SF-36 PF Score: Relationship to FVC

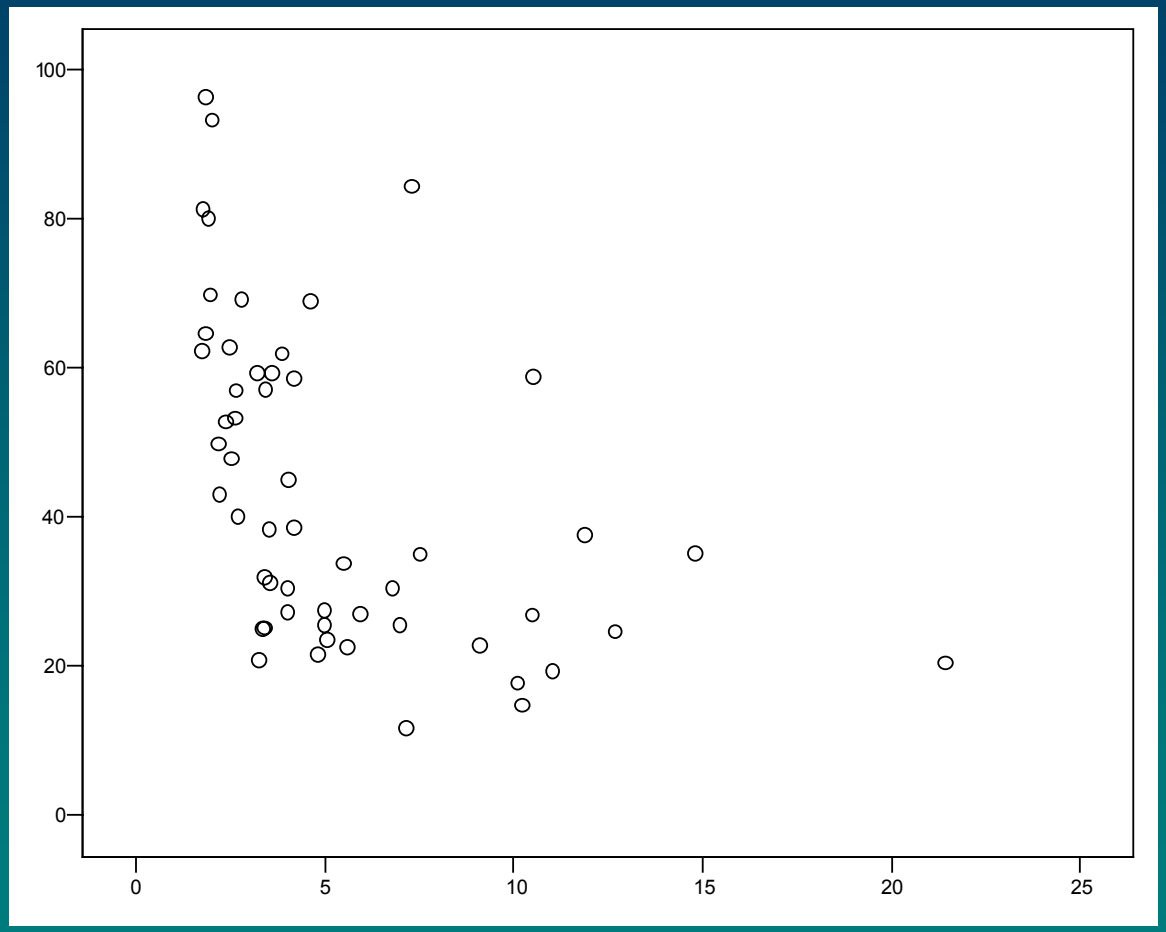


Results based on LOPOS Month 12 study visit

Lower Extremity Muscle Weakness and Timed Walk Test



Lower Extremity Muscle Weakness and Timed Stair Climb



QMT: Feasibility

- Requires equipment ($$ less than EMG machine)
- Installation is simple, and space can be shared with regular clinic room
- Time: about 1 hour of testing with cooperative patient
- Easy to learn and train evaluators and maintain reliability

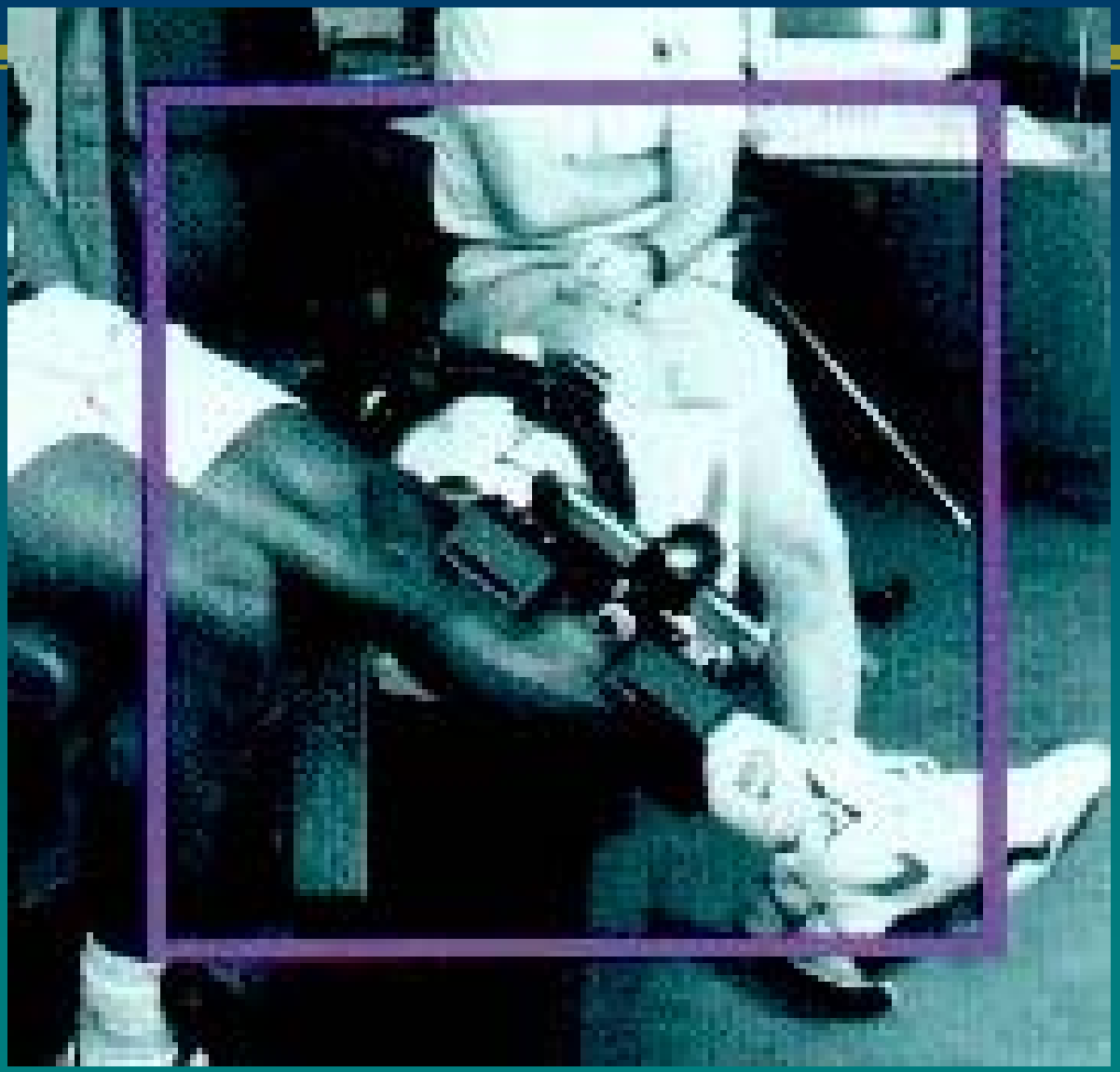
Summary of CQMS

- **Benefits**
  - Sensitive
  - Reliable
  - Can be used for individual muscle group testing
  - Requires little training
  - Reasonably validated for muscle disorders ? (DMD, FSH, Pompe, MD)
  - Quality of data improves by direct electronic capture and transmission
  - Normative data exist in adults
- **Drawbacks**
  - Expensive (CQMS, TUFT) (less than EMG machines)
  - Too sensitive to increase in muscle mass due to normal growth
  - Gaps in knowledge: what effect size is clinically meaningful in DMD, FSH and other disorders?
  - No normative data for DMD

Instrumented Systems

- “Moving systems” – Isokinetic devices
- Isometric systems

Isokinetic Dynamometer



Isokinetic Devices

- Electromechanical device with movable lever arm
- User sets velocity
- Torque is measured as patient attempts to accelerate limb past pre-set velocity
- Results in interval scale data
- Can be used for exercise

Isokinetic Devices

- Devices are bulky and expensive
- Time-intensive testing set-up
- Torque measurements probably only valid for hinge joints (knee or elbow)
- Gravity correction
  - Error amplified in low force measurements
- Have not been used in clinical trials for muscle disorders

Break Out Session: Muscle Strength

Method	reliable	sensitive	Reasonably validated	Meaningful	Feasible	Comment
MMT						
QMT						
Myometer						
Isokinetic						
Isotonic						
??						
??						

The issues

- Measure one muscle group or a total score
  - Different diseases=different needs
    - Symmetric: arm, leg, total scores correlate
    - ALS: asymmetric: MMT was better
  - Different objectives=different needs
    - Local intramuscular injection (gene therapy) needs QMT of that muscle group
    - Systemic delivery and effect expected: needs overall measure
      - Large effect size expected: less sensitive measure (i.e.MMT)
      - Small effect size expected, add-on trials: need more sensitive measure (i.e. QMT)

The issues

- 2. Reliability
  - The more reliable, the smaller the sample size
    - Children vs. adults
    - Single center vs. multi-center trials
    - Need for training
      - QMT easy to train/MMT takes longer and more experience

The issues

- 3. Sensitivity
  - Proof of principle trial vs. phase III trial
  - Monotherapy vs, add-on (effect size)
  - Measure of responsiveness

The issues

- 4. Clinical meaningfulness of changes in muscle strength
  - Correlation with function? As measured by what? TFTs? Pedometer? Other functional outcomes? Parent/patient reported outcome?
  - What change in each parameter is clinically meaningful?
    - Experience of the myositis group