On cognitive performance as endpoint in clinical trials

Ben Schmand

Department of Neurology, Academic Medical Center
Department of Psychology, University of Amsterdam
The Netherlands

Disclosure Schmand

- Royalties for test authorships
  (paid to University of Amsterdam)
  - Pearson Assessment BV Amsterdam
  - PITS Leiden / Hogrefe Publishers Amsterdam
- Royalties for book editorships
  (paid to Stichting Neuropsychologie Nederland)
  - Pearson Assessment BV Amsterdam
  - Boom Publishers Amsterdam

What are the best endpoints for clinical trials in MCI?

- FDA wants cognitive and functional measures
- ADAS-cog traditional cognitive measure
- ADAS-cog not sensitive to change in MCI
- Can neuroimaging provide better endpoints?
- Or neuropsychological assessment?

Can neuroimaging provide better endpoints?

MRI as a biomarker of disease progression in a therapeutic trial of milameline for AD

C.R. Jack, Jr., MD; M. Slomkowski, PharmD; S. Gracon, DVM; T.M. Hoover, PhD; J.P. Felmlee, PhD; K. Stewart, BS; Y. Xu, MD, PhD; M. Shiung, BA; P.C. O'Brien, PhD; R. Cha, MS; D. Knopman, MD; and R.C. Petersen, PhD, MD

Jack et al. Neurology 2003

Table 2 *Annual change from baseline in behavioral/cognitive and MRI variables*

Variable	Raw change treated (n = 100)	Raw change placebo (n = 92)	Overall median *raw* change (n = 192)	Overall median annual *percent* change (n = 192)	Percent decliners*
ADAS-Cog†	4.8 (−10.7, 25.6)	3.5 (−21.5, 19.9)	4.1 (−21.5, 25.6)	16.4 (−59.9, 152.9)	60.4
MMSE‡	−2.1 (−16.2, 6.4)	−1.1 (−18.1, 7.2)	−1.9 (−18.1, 7.2)	−8.3 (−181.1, 48.6)	66.2
GDS§	0 (−1.3, 2.4)	0 (−2.4, 2.6)	0 (−2.4, 2.6)	0.0 (−47.6, 95.4)	38.5
Total hippocampal (mm^3)	−221 (−665, 19)	−220 (−674, −7)	−221 (−674, 19)	−4.9 (−15.2, 0.5)	99.0
Total temporal horn volume (mm^3)	658 (−576, 3241)	497 (−623, 2541)	616 (−623, 3241)	16.1 (−13.1, 53.5)	85.4
Duration between 2 scans (mo)	12.3 (9, 14)	12.2 (9, 15)	12.2 (9, 15)		

Values in table represent median and range.

Table 2 *Annual change from baseline in behavioral/cognitive and MRI variables*

Variable	Raw change treated (n = 100)	Raw change placebo (n = 92)	Overall median *raw* change (n = 192)	Overall median annual *percent* change (n = 192)	Percent decliners*
ADAS-Cog†	4.8 (−10.7, 25.6)	3.5 (−21.5, 19.9)	4.1 (−21.5, 25.6)	16.4 (−59.9, 152.9)	60.4
MMSE‡	−2.1 (−16.2, 6.4)	−1.1 (−18.1, 7.2)	−1.9 (−18.1, 7.2)	−8.3 (−181.1, 48.6)	66.2
GDS§	0 (−1.3, 2.4)	0 (−2.4, 2.6)	0 (−2.4, 2.6)	0.0 (−47.6, 95.4)	38.5
Total hippocampal (mm^3)	−221 (−665, 19)	−220 (−674, −7)	−221 (−674, 19)	−4.9 (−15.2, 0.5)	99.0
Total temporal horn volume (mm^3)	658 (−576, 3241)	497 (−623, 2541)	616 (−623, 3241)	16.1 (−13.1, 53.5)	85.4
Duration between 2 scans (mo)	12.3 (9, 14)	12.2 (9, 15)	12.2 (9, 15)		

Values in table represent median and range.

Placebo controlled trial, 1 year duration,
effect size = 50% reduction in rate of change,
90% power, p<.05 one-tailed
ADAS-cog score: n=320 per arm
Hippocampal atrophy: n=21 per arm

Required sample size in RCTs

Is a function of

- Size of effect one wants to detect Δ
- Variance in untreated patients σ^2
- Level of statistical significance α
- Statistical power of the study $1-\beta$

$$n / arm = 2 (z_{1-\alpha/2} + z_{1-\beta})^2 \sigma^2 / \Delta^2$$

Journal of Alzheimer's Disease 26 (2011) 369–377
DOI 10.3233/JAD-2011-0062
IOS Press

Review

Power Calculations for Clinical Trials in Alzheimer's Disease

M. Colin Ard[a] and Steven D. Edland[a,b,*]
[a]*Department of Neuroscience, University of California, San Diego, La Jolla, CA, USA*
[b]*Department of Family Preventive Medicine Division of Biostatistics, University of California, San Diego, La Jolla, CA, USA*

Sample size required to detect a 25% reduction in annual rate of change for ADAS-Cog scores in AD and MCI (80% power and two-sided $\alpha = 0.05$)

Paper	Trial design	AD		MCI	
	Yrs (#Obs)	ROC	n/arm	ROC	n/arm
Aisen et al. (2010)[a]	1(n.s.)	4.3	407	1.1	4099
Beckett et al. (2010)[a]	2(5)	4.37	–	1.05	375
Chen et al. (2010)[a]	1(2)	3.8	[353, 505][b]	1.0	[4026, 4219][b]
Fleisher et al. (2009)[a]	1(n.s.)	n.s.	[474, 612][c]	–	–
	2(n.s.)[d]	–	–	n.s.	854
Ho et al. (2010)[a]	1(2)	3.29	583	2.43	1183
Holland et al. (2009)[e]	1(3)	4.08	624	1.19	4167
	2(5)	4.84	324	1.44	1232
Hua et al. (2010)[a]	0.5(2)	n.s.	1371	n.s.	16645
	1(2)	n.s.	483	n.s.	8212
	1.5(2)	–	–	n.s.	1381
	2(2)	n.s.	215	n.s.	1013
Landau et al. (2009)[f]	1(3)	3.8	312	1.0	2175
McEvoy et al. (2010)[e]	2(5)	–	–	1.47	978
Nestor et al. (2008)[a,d]	0.5(2)	4.8	769	1.2	9500+
Schuff et al. (2009)[d,g]	0.5(2)	n.s.	557	n.s.	3484
	1(2)	n.s.	609	n.s.	6985
	1(3)	n.s.	426	n.s.	6241

ADAS-cog is not very sensitive to change in AD, and even less in MCI

Sample size required to detect a 25% reduction in annual rate of hippocampal atrophy in AD and MCI (80% power and two-sided $\alpha = 0.05$)

Paper	Trial design	AD		MCI	
	Yrs (#Obs)	Atrophy rate	n/arm	Atrophy rate	n/arm
Aisen et al. (2010) [a]	1(n.s.)	n.s.	99	n.s.	208
Beckett et al. (2010)[a]	1(n.s.)	(116 mm^3)	–	(80 mm^3)	202
Holland et al. (2009)[b]	1(3)	3.42%	111	2.10%	235
	2(5)	3.28%	67	1.96%	179
Hua et al. (2010)[a]	0.5(2)	n.s.	114	n.s.	143
	1(2)	n.s.	68	n.s.	125
	1.5(2)	n.s.	–	n.s.	117
	2(2)	n.s.	84	n.s.	103
Leung et al. (2010)[a]	1(2)[c1]	4.57%	78	2.86%	196
	1(2)[c2]	4.58%	170	3.68%	285
McEvoy et al. (2010)[b]	2(5)	–	–	1.93%	186
Schuff et al. (2009)[d,e]	0.5(2)	3.3% (53.5 mm^3)	346	2.0% (37.7 mm^3)	709
	1(2)	4.4% (72.0 mm^3)	189	2.6% (47.5 mm^3)	522
	1(3)	n.s.	191	n.s.	503
Wolz et al. (2009)[a]	1(2)	3.85%	67	2.34%	206
	2(3)[f]	3.37%	46	2.25%	121
Yushkevich et al. (2010)[a]	1(2)	–	–	2.04%	220[g]

But what about a proper neuropsychological evaluation?

Improving the early Diagnosis of Alzheimer's Disease and Other dementias (IDADO study)

- Memory clinic patients
- Inclusion criteria:
  - Possibly in early stage of dementia
  - Baseline and follow-up NP assessment + MRI scan
- Exclusion criteria:
  - Dementia at baseline (clinical diagnosis)
  - Non-credible responding during NP assessment
  - Other brain disease that explains symptoms

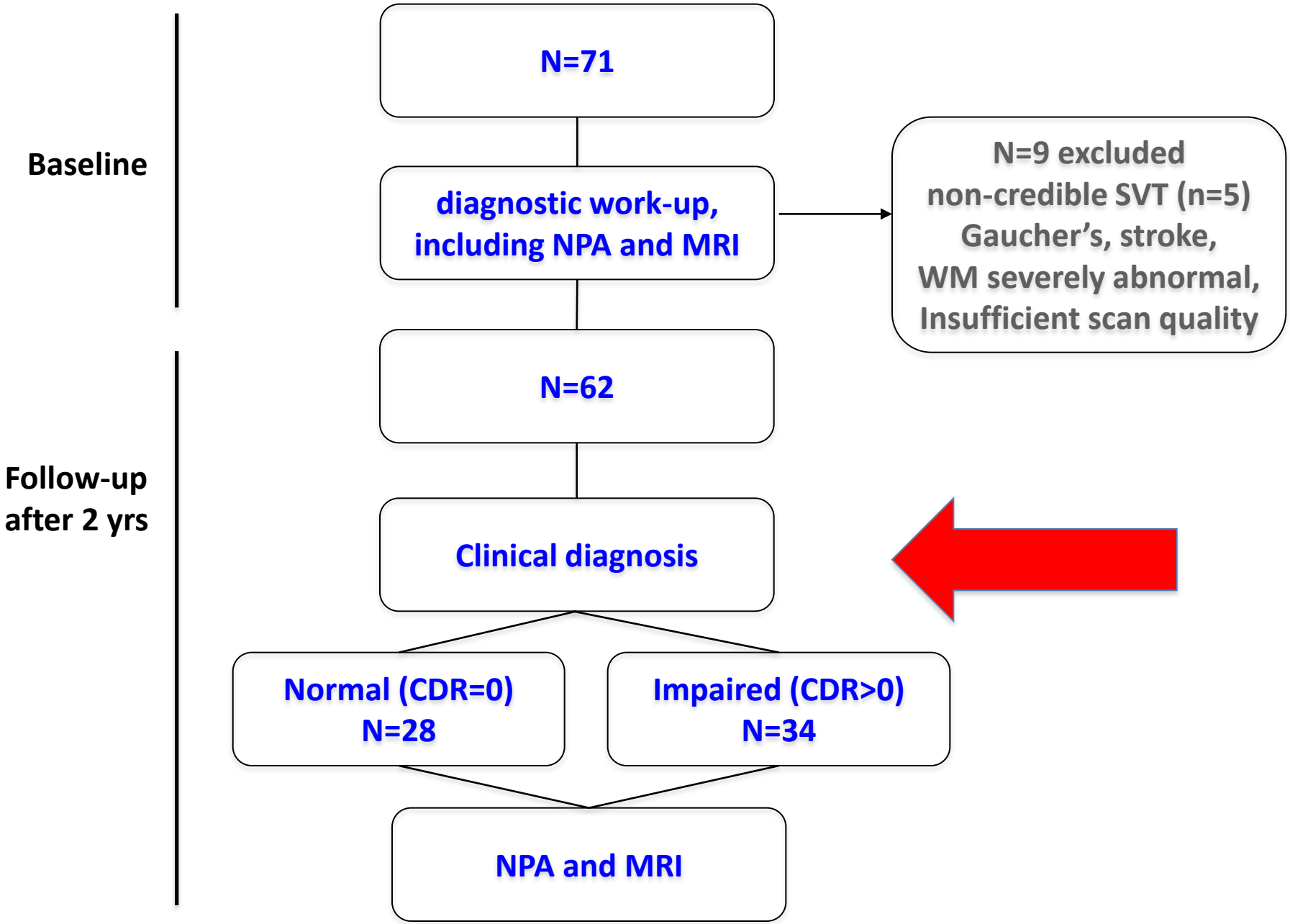
IDADO study collaboration

- Anne Rienstra
- Hyke Tamminga
- Edo Richard
- Willem A. van Gool
- Gerard Walstra
- Nikki Lammers
- Ben Schmand
- Matthan Caan
- Charles B. Majoie
- ***Neurology & Radiology, Academic Medical Center Psychology department, University of Amsterdam***

- Jos de Jonghe
- ***Medical Center Alkmaar***

- Ton d'Hondt
- ***GGZ Noord Holland***

- Bregje Appels
- Jos van Campen
- ***Slotervaart Hospital***


Baseline
N=71
diagnostic work-up, including NPA and MRI
N=9 excluded
non-credible SVT (n=5)
Gaucher's, stroke,
WM severely abnormal,
Insufficient scan quality
Follow-up after 2 yrs
N=62
Clinical diagnosis
Normal (CDR=0)
N=28
Impaired (CDR>0)
N=34
NPA and MRI

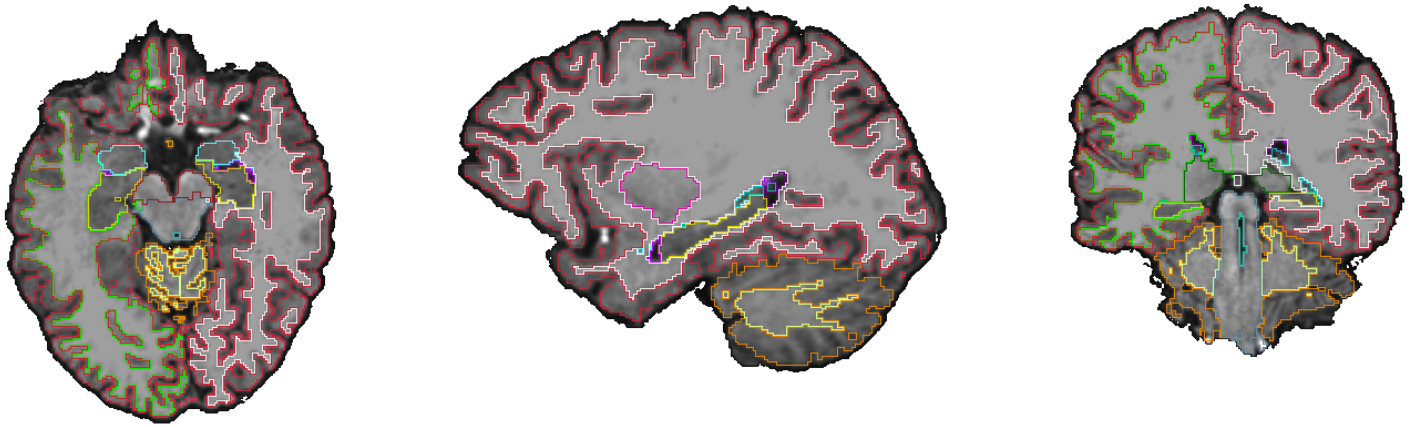
Patient characteristics at baseline and follow-up

	Normal cognition at follow-up (CDR=0; n=28)	**Cognitive impairment at follow-up** (CDR>0; n=34)	p-value
Male / female	14 / 14	18 / 16	.82
Age	62.0 (8.1)	70.3 (8.9)	<.001
Education level (0-6 ISCED)	4.1 (1.4)	3.8 (1.2)	.27
MMSE at baseline	28.0 (1.8)	26.7 (2.3)	.011
MMSE at follow-up	28.2 (1.3)	23.0 (4.5)	<.001

Neuropsychological tests

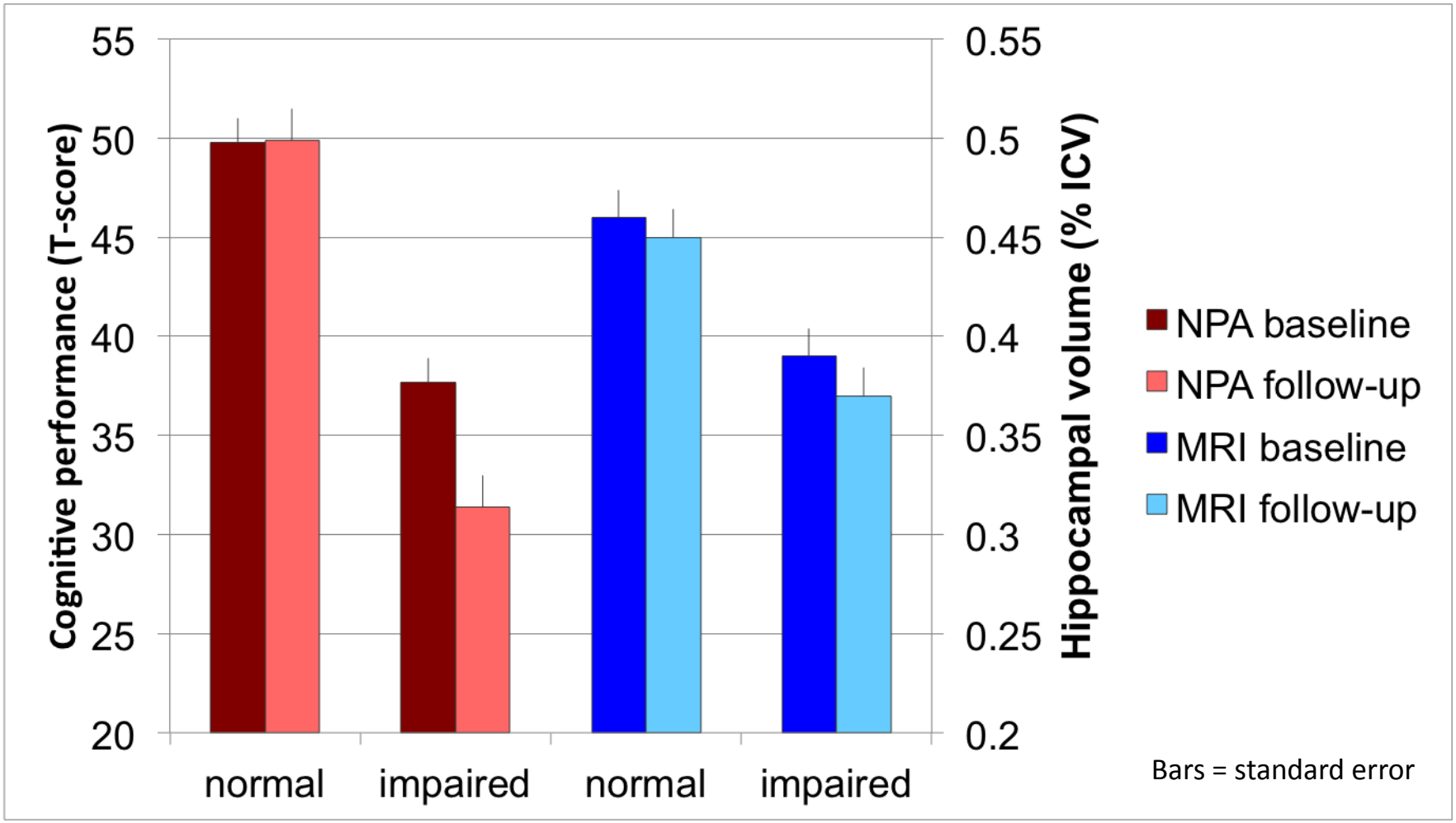
- Rey's AVLT immediate recall
- Rey's AVLT delayed recall
- Rivermead BMT prose immediate recall
- Rivermead BMT prose delayed recall
- Letter fluency (COWAT)
- Stroop Color Word Test interference
- Trail Making Test part B
- T-scores are age, gender & education corrected
  Normally distributed in the general population
  T-scores: mean = 50, standard deviation = 10

FreeSurfer automatic partitioning and volumetry (3 Tesla MRI)



MRI measures:
hippocampal volume as percentage of intracranial volume
and
cortical thickness of
entorhinal, middle temporal, and parahippocampal areas

Cognitive performance (L) and hippocampal volume (R)
patients with normal cognition and declining patients



N needed per arm

$$n / arm = 2 (z_{1-\alpha/2} + z_{1-\beta})^2 \sigma^2 / \Delta^2$$

Placebo controlled trial, effect size 50% reduction in rate of change, 80% power, $p<.05$ one-tailed

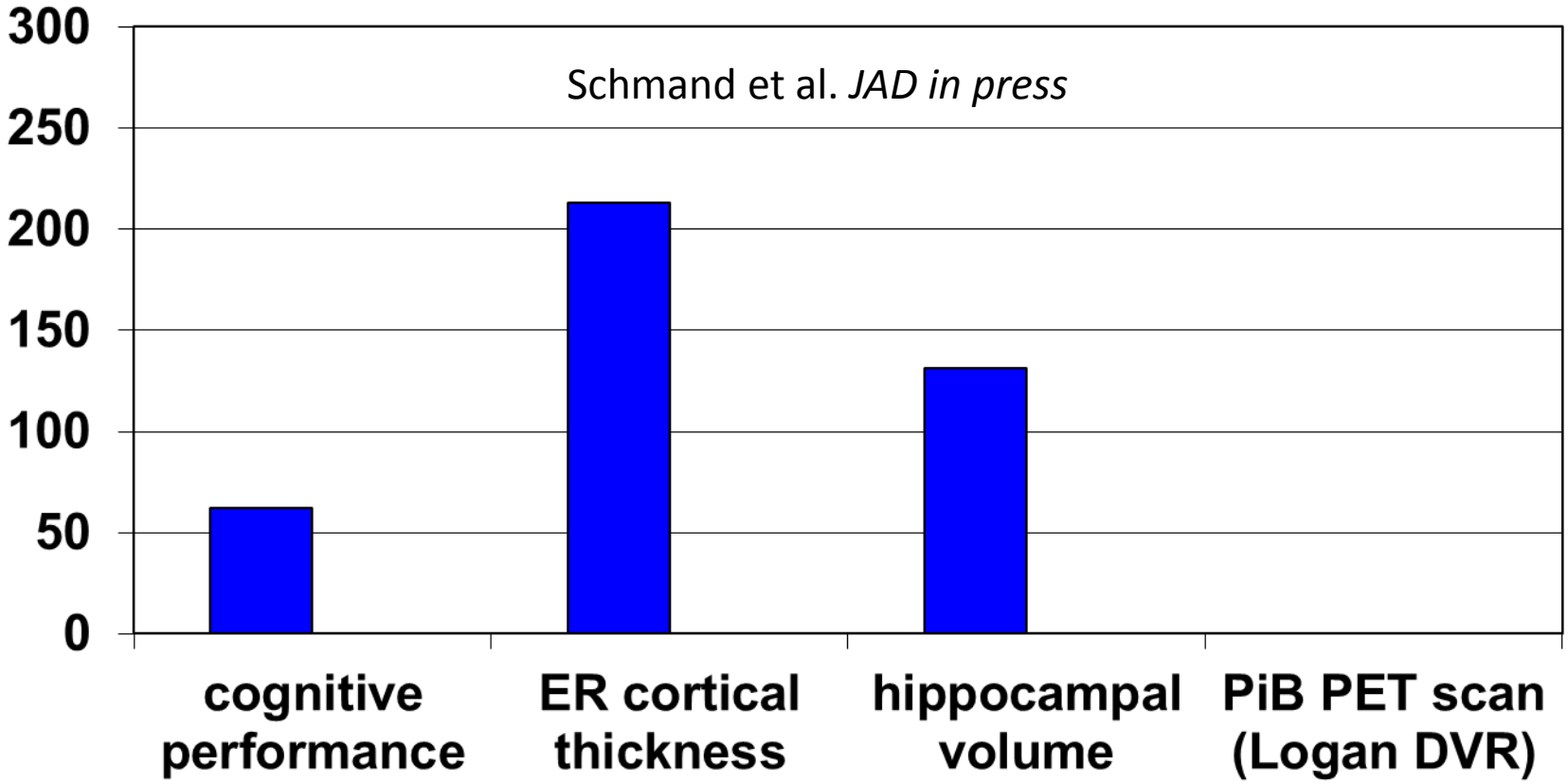
Hippocampal atrophy: n=131 per arm
Neuropsychological tests: n=62 per arm

Note: Δ = 50% of (mean change $_{impaired}$ – mean change $_{normal}$)
Thus delta is corrected for change in normal group

523 and 246 per arm for 25% reduction in rate of change

N needed per arm for various outcomes in a hypothetical RCT

to detect 50% reduction in rate of change at 80% study power



Longitudinal Amyloid Imaging Using ^{11}C-PiB: Methodologic Considerations

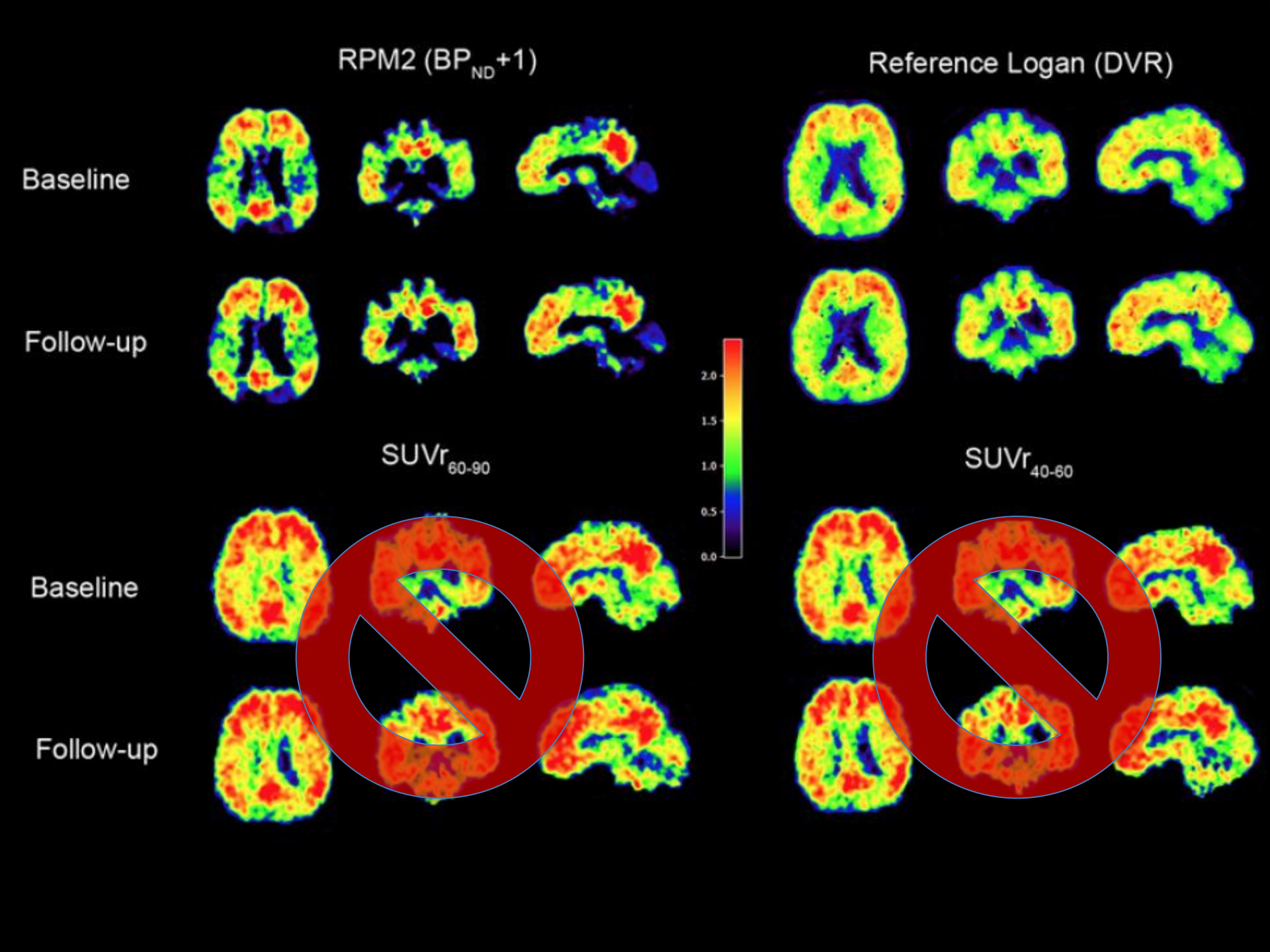
Bart N.M. van Berckel[1], Rik Ossenkoppele[1,2], Nelleke Tolboom[1,2], Maqsood Yaqub[1], Jessica C. Foster-Dingley[1], Albert D. Windhorst[1], Philip Scheltens[2], Adriaan A. Lammertsma[1], and Ronald Boellaard[1]

[1]Department of Nuclear Medicine and PET Research, VU University Medical Center, Amsterdam, The Netherlands; and [2]Department of Neurology and Alzheimer Center, VU University Medical Center, Amsterdam, The Netherlands

Van Berckel et al. J Nucl Med 2013

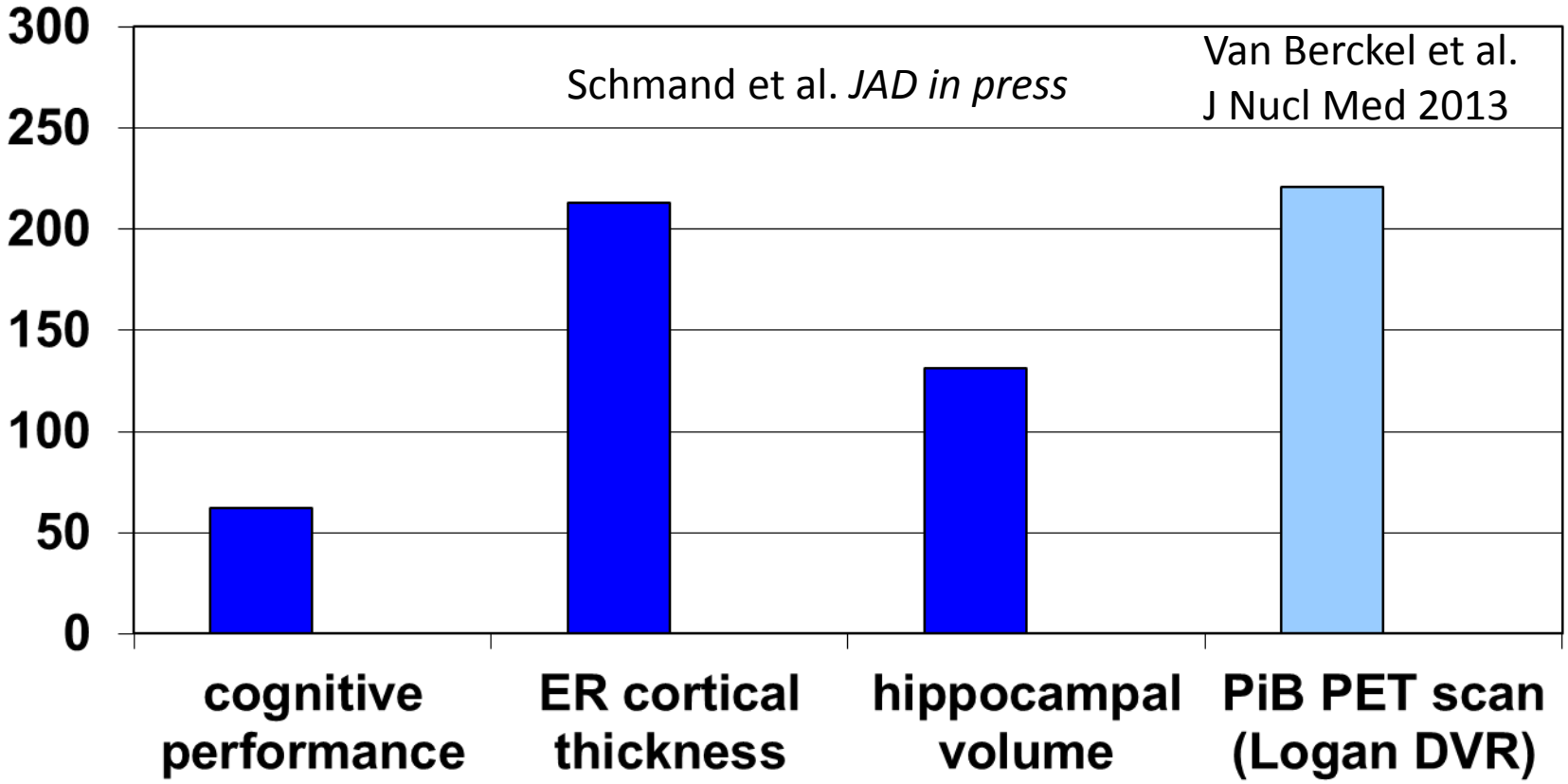
MCI-patients (n=11) and controls (n=11)

PiB PET scanning at baseline and after 2.5 years
comparison of four analytic techniques


RPM2 (BP$_{ND}$+1)
Reference Logan (DVR)
Baseline
Follow-up
2.0
1.5
1.0
0.5
0.0
SUVr$_{60-90}$
SUVr$_{40-60}$
Baseline
Follow-up

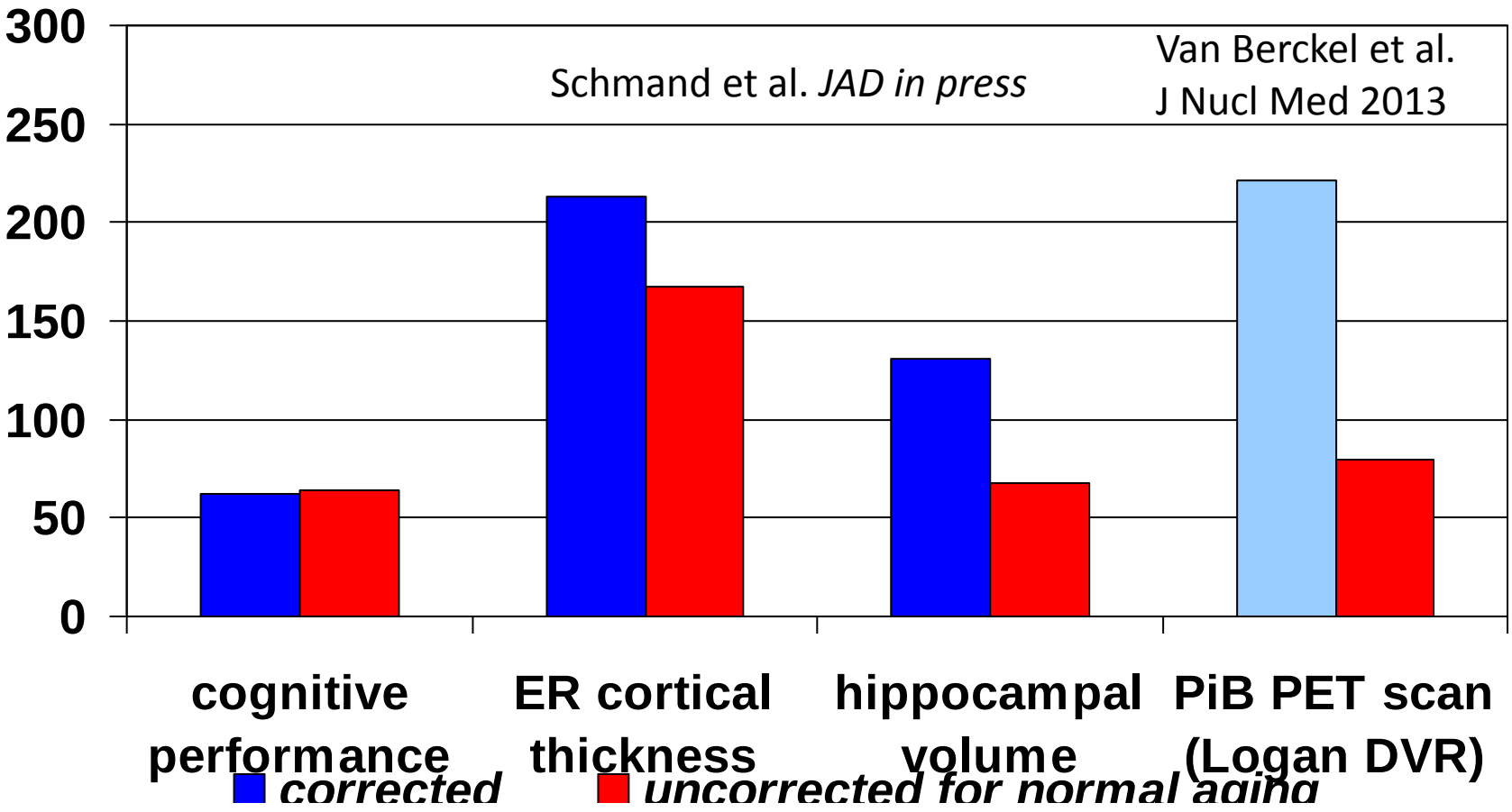
N needed per arm for various outcomes in a hypothetical RCT

to detect 50% reduction in rate of change at 80% study power



N needed per arm for various outcomes in a hypothetical RCT

to detect 50% reduction in rate of change at 80% study power



What are the best endpoints for clinical trials in MCI?

- FDA wants cognitive and functional measures
- ADAS-cog traditional cognitive measure
- ADAS-cog not sensitive to change in MCI
- Can neuroimaging provide better endpoints?
- Or neuropsychological assessment?
- FDA prepared to consider NP assessment?
  (Draft Guidance for Industry, February 2013)

Bottom line & take home message

- **Track disease course or evaluate treatment? Then stick to the symptoms! (axiom)**
- **Cognitive performance is most sensitive to change in MCI**
- **Cognition should remain a primary endpoint provided it is measured in a sound way**


b.schmand@amc.nl