Adaptive Clinical Trials: A Necessary Step toward Personalized Medicine

Donald A. Berry
dberry@mdanderson.org

THE UNIVERSITY OF TEXAS
MD Anderson Cancer Center
Making Cancer History®

Berry Consultants
Statistical Innovation

Financial Disclosure

- Part owner Berry Consultants, LLC
- Designs adaptive trials for
  - Pharmaceutical companies
  - Medical device companies
  - NIH-sponsored cooperative groups

Janet Woodcock, Director CDER FDA

“Improved utilization of adaptive and Bayesian methods” could help resolve low success rate of and expense of phase 3 clinical trials

FDA's Critical Path Opportunities Report (2006)

"uncovered a consensus that the two most important areas for improving medical product development are **biomarker development** and **streamlining clinical trials**."

http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/default.htm

The NEW ENGLAND JOURNAL of MEDICINE

Perspective

Development of Novel Combination Therapies

Janet Woodcock, M.D., Joseph P. Griffin, J.D., and Rachel E. Behrman, M.D., M.P.H.

For example, in 2010, the Biomarkers Consortium–a public-private partnership that includes the NIH, the FDA, patient groups, and pharmaceutical and biotech–initiated a groundbreaking trial in breast cancer to predict drug responsiveness based on the presence or absence of genetic and biological markers, … I-SPY 2 (ClinicalTrials.gov NCT01042379).

This article (10.1056/NEJMp1101548) was published on February 16, 2011, at NEJM.org.

cific molecules, inclu… gram. Increasingly, tumors will contributing to the pr… screened for pertinent path- of cancer cells and … dependencies, as is current- microorganisms. Although target- …tively treat many tumors and in- …ly done for breast cancer, and pa-

ScienceInsider

Breaking news and analysis from the world of science policy

FDA's $25 Million Pitch for Improving Drug Regulation

by Jennifer Couzin-Frankel on 7 October 2010, 3:17 PM | Permanent Link | 0 Comments

"Our current approach [to trials] is horribly inefficient, and we need to do something better," says Roger Lewis, an emergency medicine physician at Harbor-University of California, Los Angeles, Medical Center. Lewis helps advise a company called Berry Consultants …

FDA is trying to move forward nevertheless, in part by linking up with more flush agencies. Last week, in conjunction with the National Institutes of Health (NIH), it announced four sizable grants, totaling $9.4 million, in regulatory science. (FDA contributed just under $1 million and NIH gave the rest.) They include support for a heart-lung system that can test potential drugs and an effort to dramatically streamline clinical trials.

"Our current approach [to trials] is horribly inefficient, and we need to do something better," says Roger Lewis, an emergency medicine physician at Harbor-University of California, Los Angeles, Medical Center. Lewis helps advise a company called Berry Consultants founded by Donald Berry, a biostatistician at M.D. Anderson Cancer Center in Houston, Texas. He and Berry, along with emergency medicine physician William Barsan at the

ScienceInsider

Breaking news and analysis from the world of science policy

FDA's $25 Million Pitch for Improving Drug Regulation

by Jennifer Couzin-Frankel on 7 October 2010, 3:17 PM | Permanent Link | 0 Comments

Lewis and Berry, along with emergency medicine physician William Barsan at the University of Michigan, will be studying whether "adaptive" trial designs that incorporate new information in midcourse can answer medical questions. They also want to learn what concerns researchers might have about this approach.

conjunction with the National Institutes of Health (NIH), it announced four sizable grants, totaling $9.4 million, in regulatory science. (FDA contributed just under $1 million and NIH gave the rest.) They include support for a heart-lung system that can test potential drugs and an effort to dramatically streamline clinical trials.

"Our current approach [to trials] is horribly inefficient, and we need to do something better," says Roger Lewis, an emergency medicine physician at Harbor-University of California, Los Angeles, Medical Center. Lewis helps advise a company called Berry Consultants founded by Donald Berry, a biostatistician at M.D. Anderson Cancer Center in Houston, Texas. He and Berry, along with emergency medicine physician William Barsan at the

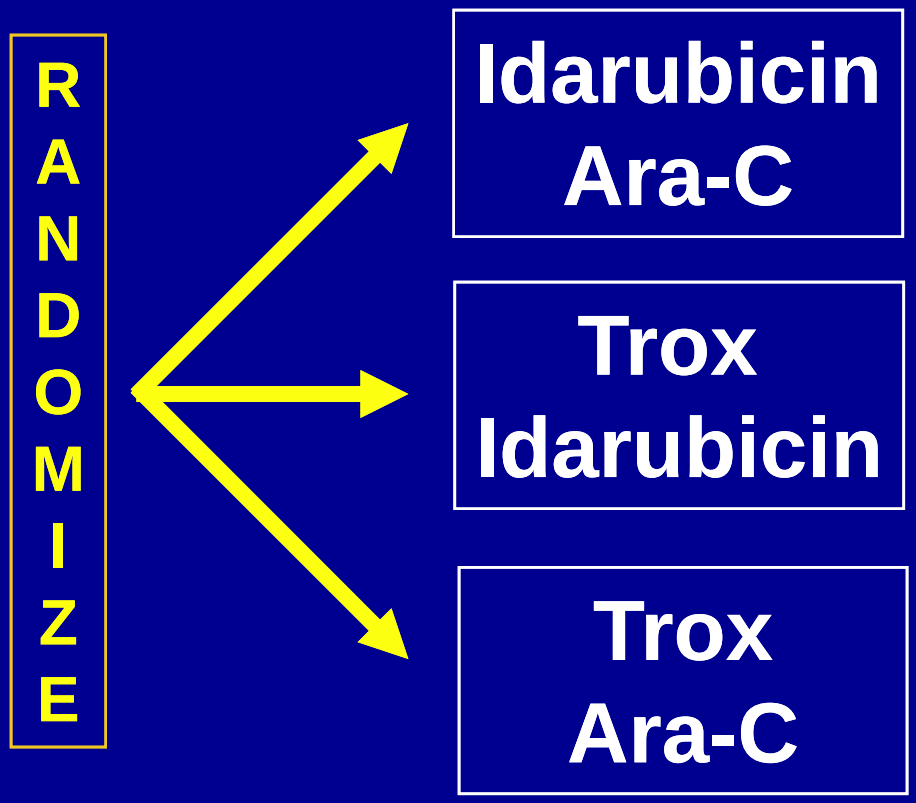
Current use of Bayesian adaptive designs

- MDACC (> 300 trials)
- Device companies (> 25 PMAs)*
- Drug companies (Most of top 40; many biotechs)**

*http://www.fda.gov/MedicalDevicesDeviceRegulationandGuidance/GuidanceDocuments/ucm071072.htm

**http://www.fda.gov/downloads/DrugsGuidanceComplianceRegulatoryInformation/Guidances/UCM201790.pdf

Example: Troxacitabine in AML* (endpoint: CR by day 50)



Standard design

Adaptive randomization to learn, while effectively treating patients in trial

n = 25

n = 25

n = 25

* Giles JCO 2003

Adaptive Randomization

- Assign with higher probability to better performing therapies
- TI dropped after 24th patient
- Trial stopped after 34 patients

Summary of AML trial results

CR by 50 days:

IA	10/18 = 56%
TA	3/11 = 27%
TI	0/5 = 0%

Some areas of application of Bayesian adaptive drug trials

- Migraine
- Oncology
- Rh Arthritis
- Lupus
- Diabetes
- Obesity
- Stroke
- Gastroparesis
- Spinal Cord Injury
- HIV
- Hepatitis C
- Pre-term labor
- Constipation
- Overactive bladder
- Libido
- Alzheimer's

Simulations Usually Required

- To find operating characteristics:
  - Type I error rate
  - Power
  - Sample size distribution
- Prospective design essential
- Longitudinal modeling
- Many scenarios
- Accrual rate matters

ScienceInsider

Breaking news and analysis from the world of science policy

FDA's $25 Million Pitch for Improving Drug Regulation

by Jennifer Couzin-Frankel on 7 October 2010, 3:17 PM | Permanent Link | 0 Comments

"Our current approach [to trials] is horribly inefficient, and we need to do something better," says Roger Lewis, an emergency medicine physician at Harbor-University of California, Los Angeles, Medical Center. Lewis helps advise a company called Berry Consultants …

conjunction with the National Institutes of Health (NIH), it announced four sizable grants, totaling $9.4 million, in regulatory science. (FDA contributed just under $1 million and NIH gave the rest.) They include support for a heart-lung system that can test potential drugs and an effort to dramatically streamline clinical trials.

"Our current approach [to trials] is horribly inefficient, and we need to do something better," says Roger Lewis, an emergency medicine physician at Harbor-University of California, Los Angeles, Medical Center. Lewis helps advise a company called Berry Consultants founded by Donald Berry, a biostatistician at M.D. Anderson Cancer Center in Houston, Texas. He and Berry, along with emergency medicine physician William Barsan at the

ScienceInsider

Breaking news and analysis from the world of science policy

FDA's $25 Million Pitch for Improving Drug Regulation

by Jennifer Couzin-Frankel on 7 October 2010, 3:17 PM | Permanent Link | 0 Comments

Lewis and Berry, along with emergency medicine physician William Barsan at the University of Michigan, will be studying whether "adaptive" trial designs that incorporate new information in midcourse can answer medical questions. They also want to learn what concerns researchers might have about this approach.

regulatory science. (FDA contributed just under $1 million and NIH gave the rest.) They include support for a heart-lung system that can test potential drugs and an effort to dramatically streamline clinical trials.

"Our current approach [to trials] is horribly inefficient, and we need to do something better," says Roger Lewis, an emergency medicine physician at Harbor-University of California, Los Angeles, Medical Center. Lewis helps advise a company called Berry Consultants founded by Donald Berry, a biostatistician at M.D. Anderson Cancer Center in Houston, Texas. He and Berry, along with emergency medicine physician William Barsan at the

Bayesian adaptive trials

- Stopping early (or late)
  - Efficacy
  - Futility
- Dose finding (& dose dropping)
- Seamless phases
- Population finding
- Adaptive randomization
- Ramping up accrual

Why?

- Smaller trials (usually!)
- More accurate conclusions
- Can focus on better treatment of patients in trials

Three Recent Examples of Smaller Sample Size Using

- Bayesian predictive probabilities
- Longitudinal modeling

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 MAY 14, 2009 VOL. 360 NO. 20

Adjuvant Chemotherapy in Older Women with Early-Stage Breast Cancer

Hyman B. Muss, M.D., Donald A. Berry, Ph.D., Constance T. Cirrincione, M.S., Maria Theodoulou, M.D., Ann M. Mauer, M.D., Alice B. Kornblith, Ph.D., Ann H. Partridge, M.D., M.P.H., Lynn G. Dressler, Ph.D., Harvey J. Cohen, M.D., Heather P. Becker, Patricia A. Kartcheske, B.S., Judith D. Wheeler, M.P.H., Edith A. Perez, M.D.,

A Bayesian statistical design was used with a range in sample size from 600 to 1800 patients.

BACKGROUND

Older women with breast cancer are underrepresented in clinical trials, and data on the effects of adjuvant chemotherapy in such patients are scant. We tested for the noninferiority of capecitabine as compared with standard chemotherapy in women with breast cancer who were 65 years of age or older.

METHODS

We randomly assigned patients with stage I, II, IIIA, or IIIB breast cancer to standard chemotherapy (either cyclophosphamide, methotrexate, and fluorouracil or cyclophosphamide plus doxorubicin) or capecitabine. Endocrine therapy was recommended after chemotherapy in patients with hormone-receptor–positive tumors. A Bayesian statistical design was used with a range in sample size from 600 to 1800 patients. The primary end point was relapse-free survival.

From the University of Vermont, Burlington (H.B.M.); the M.D. Anderson Cancer Center, Houston (D.A.B.); the Cancer and Leukemia Group B (CALGB) Statistical Center, Duke University Medical Center (C.T.C., P.A.K.) and Duke University Medical Center (H.J.C., J.D.W., A.A.M.) — both in Durham, NC; Memorial Sloan-Kettering Cancer Center, New York (M.T., L.N., C.A.H.); CALGB, Chicago (A.M.M., H.P.B.); the Dana–Farber Cancer Institute, Boston (A.B.K., A.H.P., H.J.B., E.P.W.); the University of North Carolina, Chapel Hill (L.G.D.); the North Central Cancer Treat-

From “Methods”

“These interim analyses were not the standard type in which the trial results are announced when a boundary is crossed. Rather, the decision to discontinue enrollment was based on a prediction that future follow-up was likely to give a meaningful answer.”

ORIGINAL CONTRIBUTION

JAMA. 2010;303(4):333-340

Comparison of Antiarrhythmic Drug Therapy and Radiofrequency Catheter Ablation in Patients With Paroxysmal Atrial Fibrillation

A Randomized Controlled Trial

David J. Wilber, MD
Carlo Pappone, MD, PhD
Petr Neuzil, MD
Angelo De Paola, MD
Frank Marchlinski, MD
Andrea Natale, MD
Laurent Macle, MD
Emile G. Daoud, MD
Hugh Calkins, MD
Burr Hall, MD
Vivek Reddy, MD
Christine Y. Liu, MPH
Scott M. Berry, PhD
Donald A. Berry, PhD
for the ThermoCool AF Trial Investigators

Context Antiarrhythmic drugs are commonly used for prevention of recurrent atrial fibrillation (AF) despite inconsistent efficacy and frequent adverse effects. Catheter ablation has been proposed as an alternative treatment for paroxysmal AF.

Objective To determine the efficacy of catheter ablation compared with antiarrhythmic drug therapy (ADT) in treating symptomatic paroxysmal AF.

Design, Setting, and Participants A prospective, multicenter, randomized (2:1), unblinded, Bayesian-designed study conducted at 19 hospitals of 167 patients who did not respond to at least 1 antiarrhythmic drug and who experienced at least 3 AF episodes within 6 months before randomization. Enrollment occurred between October 25, 2004, and October 11, 2007, with the last follow-up on January 19, 2009.

Intervention Catheter ablation (n=106) or ADT (n=61), with assessment for effectiveness in a comparable 9-month follow-up period.

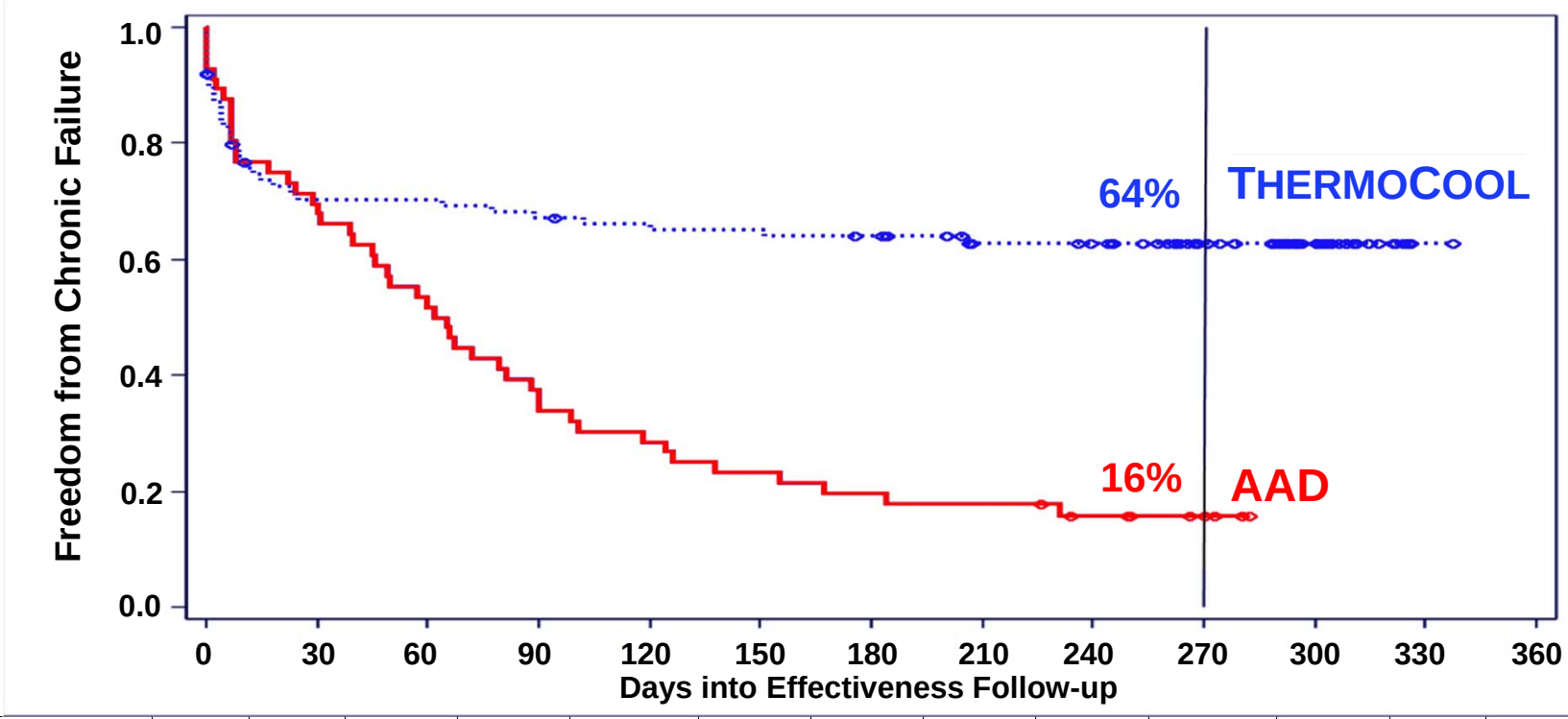
Main Outcome Measures Time to protocol-defined treatment failure. The pro-

compared with 16% of patients treated with ADT. The hazard ratio of catheter ablation to ADT was 0.30 (95% confidence interval, 0.19-0.47; P• .001). Major 30-day treatment-related adverse events occurred in 5 of 57 patients (8.8%) treated with ADT and 5 of 103 patients (4.9%) treated with catheter ablation. Mean quality of life scores improved significantly in patients treated by catheter ablation compared with ADT at 3 months; improvement was maintained during the course of the study.

Conclusion Among patients with paroxysmal AF who had not responded to at least

Design, Setting, and Participants A prospective, multicenter, randomized (2:1), unblinded, Bayesian-designed study conducted at 19 hospitals of 167 patients who

Time to Chronic Failure by Randomization Group (Updated)



TCool	103	69	69	66	63	62	61	54	52	37	15	3	2
AAD	56	39	29	19	16	13	11	10	7	2	0	0	0

ISCTM
THE INTERNATIONAL SOCIETY FOR CNS CLINICAL TRIALS AND METHODOLOGY

Adaptive Design-Key Features

Uses accumulating data to decide on how to modify aspects of the study without undermining the validity and integrity of the trial.

Response-Adaptive Randomization

Uses interim data and Bayesian techniques to modify the treatment allocation probabilities in favor of more informative treatment arms.

- Maximize information about dose-response relationship
- increases the probability of success in Phase 3 by bringing the correct dose forward into Phase 3

Predefined Decision Rules

Frequent interim analyses allows determination of futility or success

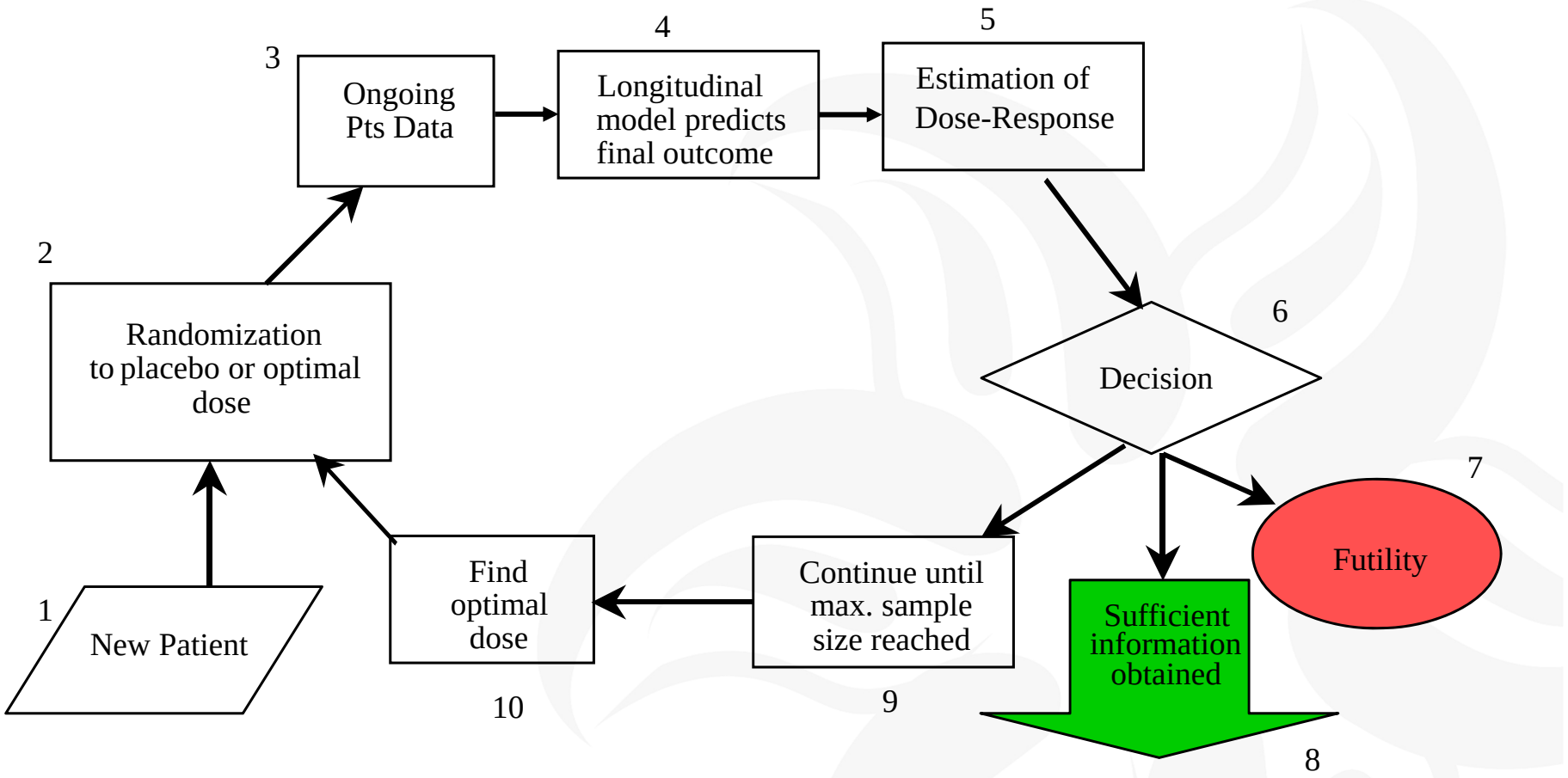
- Allows Ph2a and Ph2b to be combined in a single trial
- Succeed efficiently and fail efficiently

Trial simulations

Define operating characteristics (false + rate, statistical power, and sample size distributions)

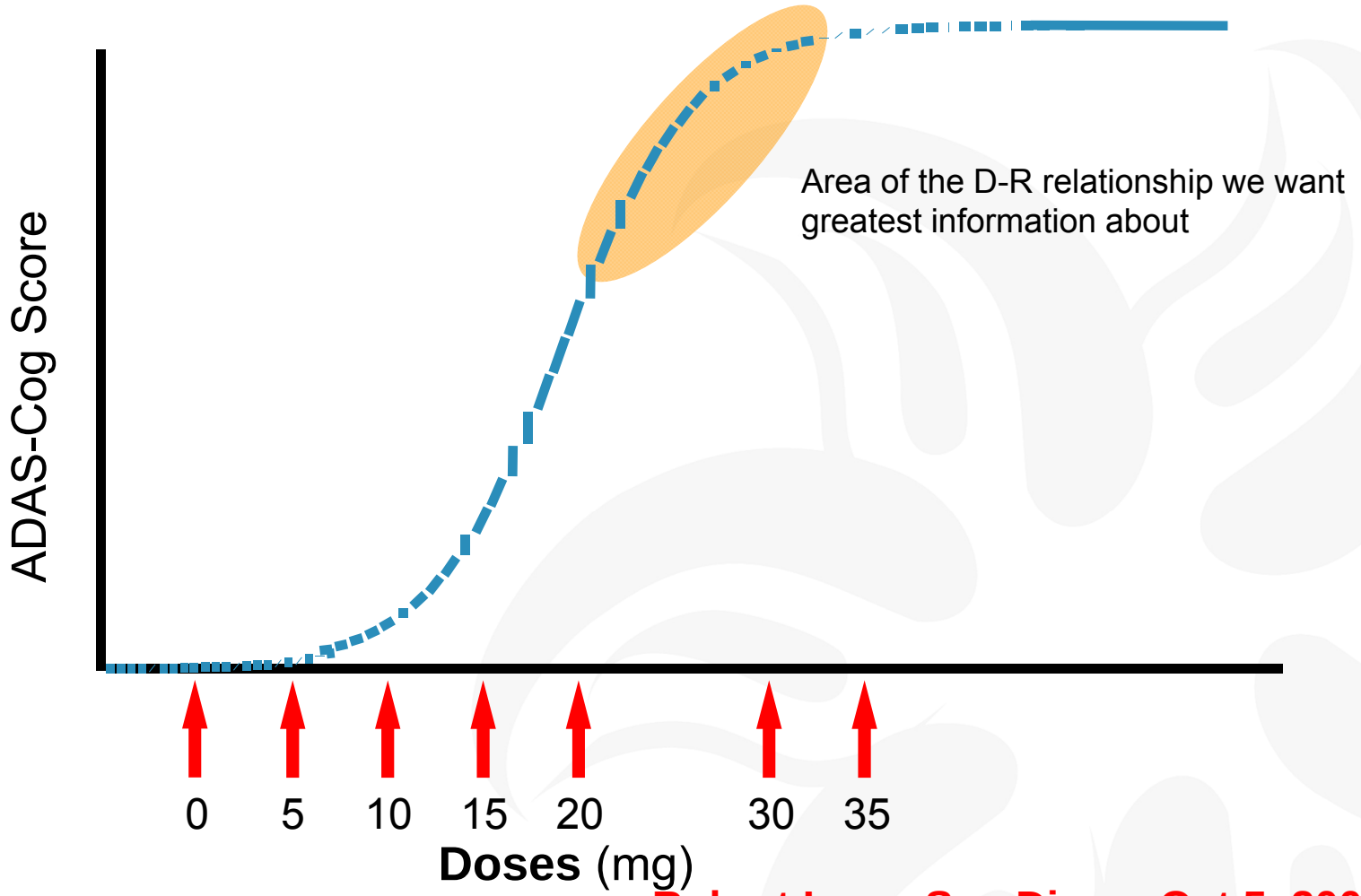
ISCTM
THE INTERNATIONAL SOCIETY FOR CNS CLINICAL TRIALS AND METHODOLOGY

Response-Adaptive Randomization



ISCTM
THE INTERNATIONAL SOCIETY FOR CNS CLINICAL TRIALS AND METHODOLOGY

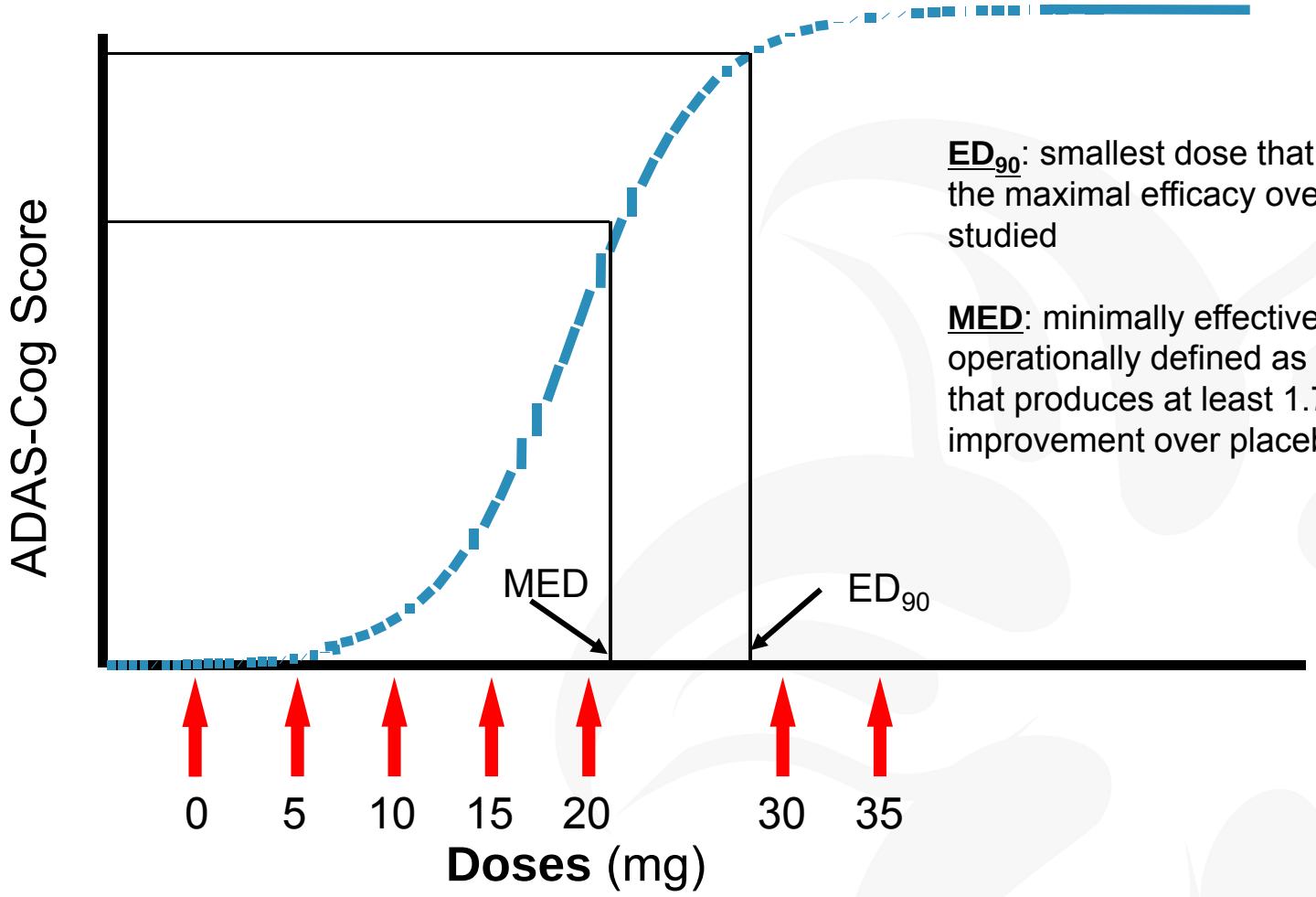
Response-Adaptive Randomization



Robert Lenz, San Diego, Oct 5, 2009

ISCTM
THE INTERNATIONAL SOCIETY FOR CNS CLINICAL TRIALS AND METHODOLOGY

Goal is to Find the ED_{90} and MED



ED_{90}: smallest dose that produces 90% of the maximal efficacy over the doses studied

MED: minimally effective dose; operationally defined as the smallest dose that produces at least 1.75 point improvement over placebo in ADAS-Cog

ISCTM
THE INTERNATIONAL SOCIETY FOR CNS CLINICAL TRIALS AND METHODOLOGY

Adaptive Randomization

- The probability of allocating the subsequent subject to a particular dose is proportional to the reduction in the variance of ED_{90} achieved when a subject is allocated to that dose.
- The allocation ratio will be altered in such as way as to maximize the information (minimize the variance) about the MED and ED_{90}.
- Randomization probabilities are updated every 2 weeks.

ISCTM
THE INTERNATIONAL SOCIETY FOR CNS CLINICAL TRIALS AND METHODOLOGY

Normal Dynamic Linear Model

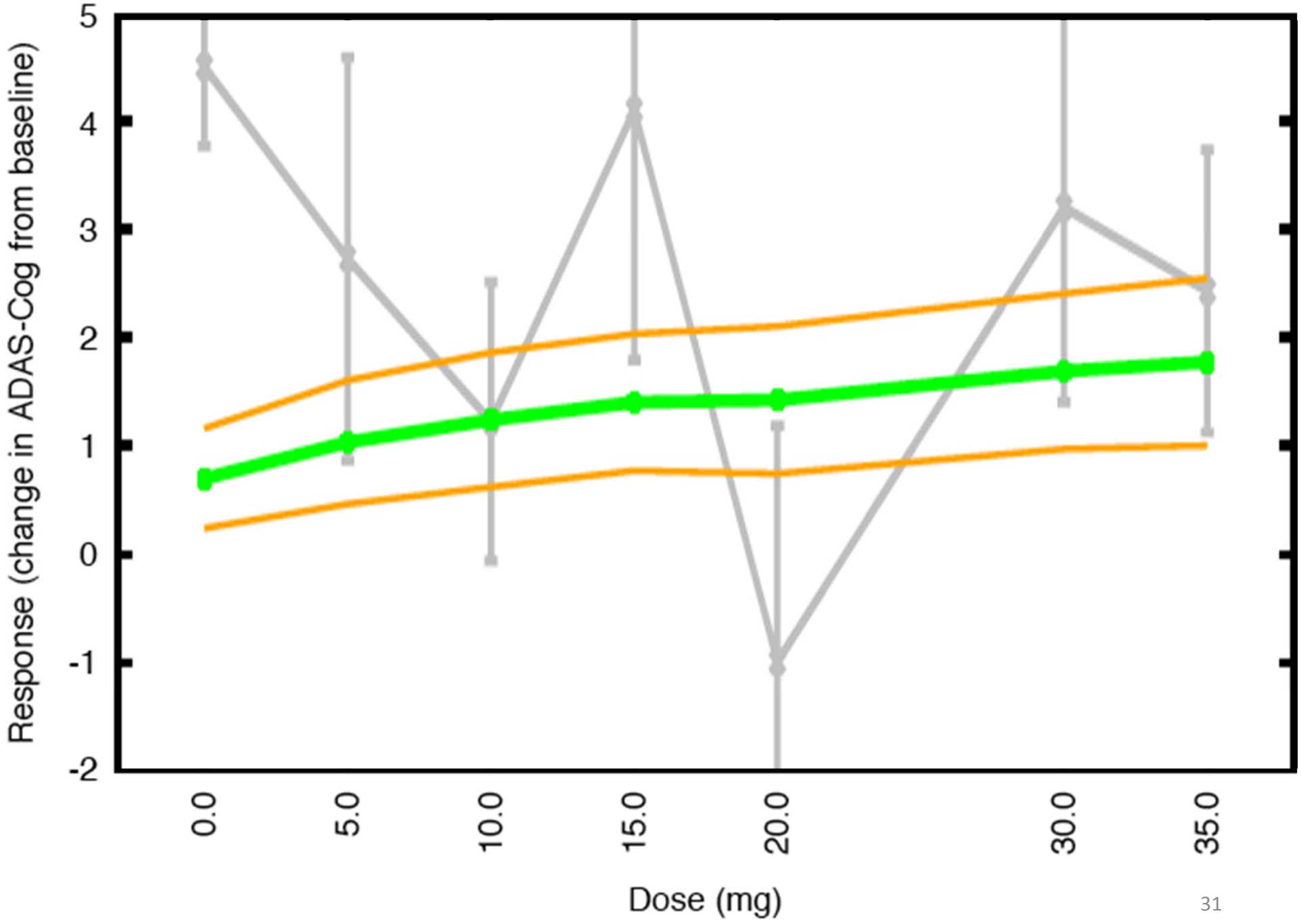
- Normal Dynamic Linear Model is used to describe the dose-response relationship.
- This is a very versatile and flexible model that allows for non-monotonic response functions.
- Allows borrowing strength from neighboring doses (don't treat each dose as entirely independent), thereby narrowing the probability distribution around the dose.
- Thus, increasing the number of doses increases information about the entire dose-response curve, not just a single dose.

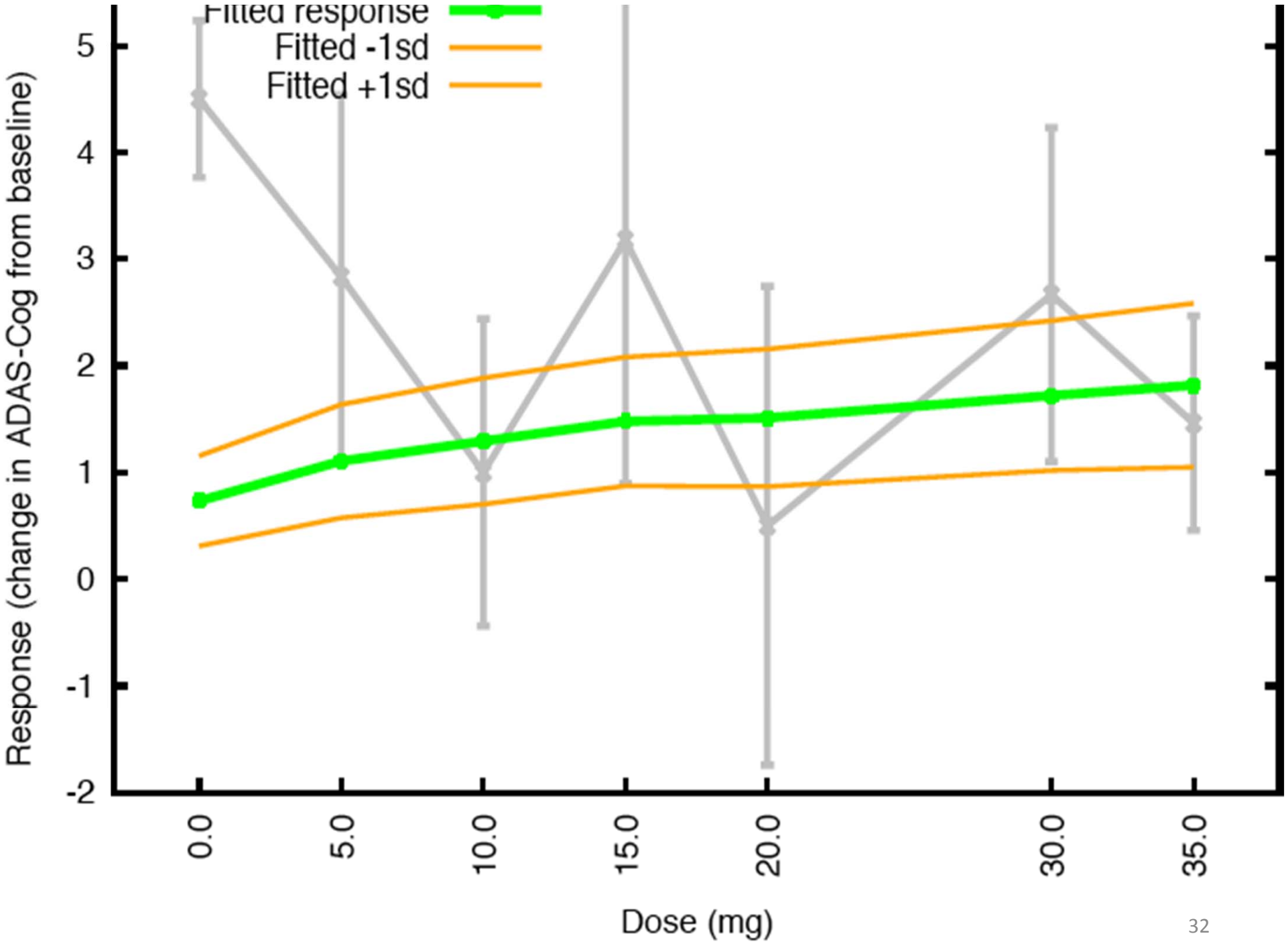
ISCTM
THE INTERNATIONAL SOCIETY FOR CNS CLINICAL TRIALS AND METHODOLOGY

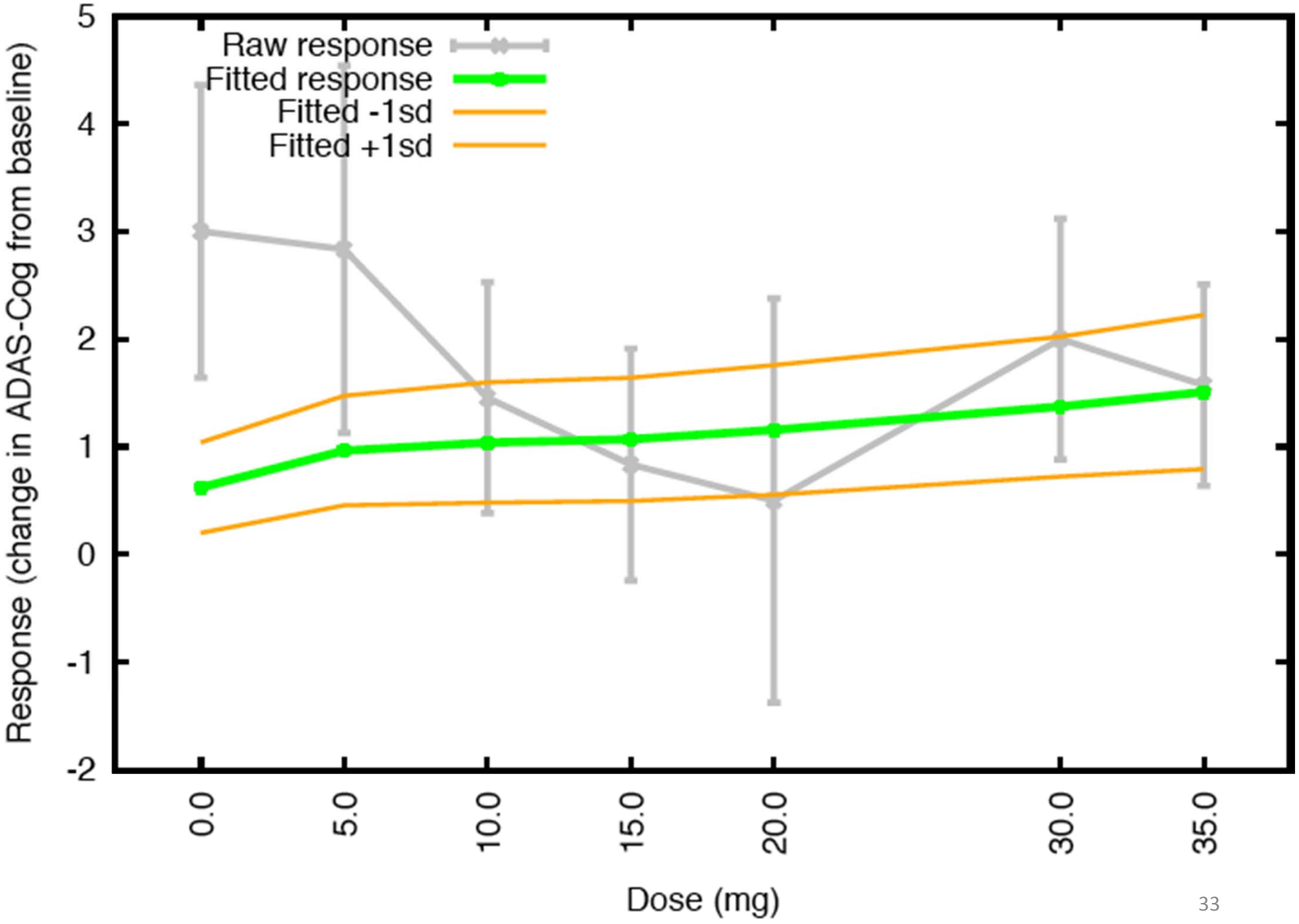
Simulations: Sensitivity Analyses

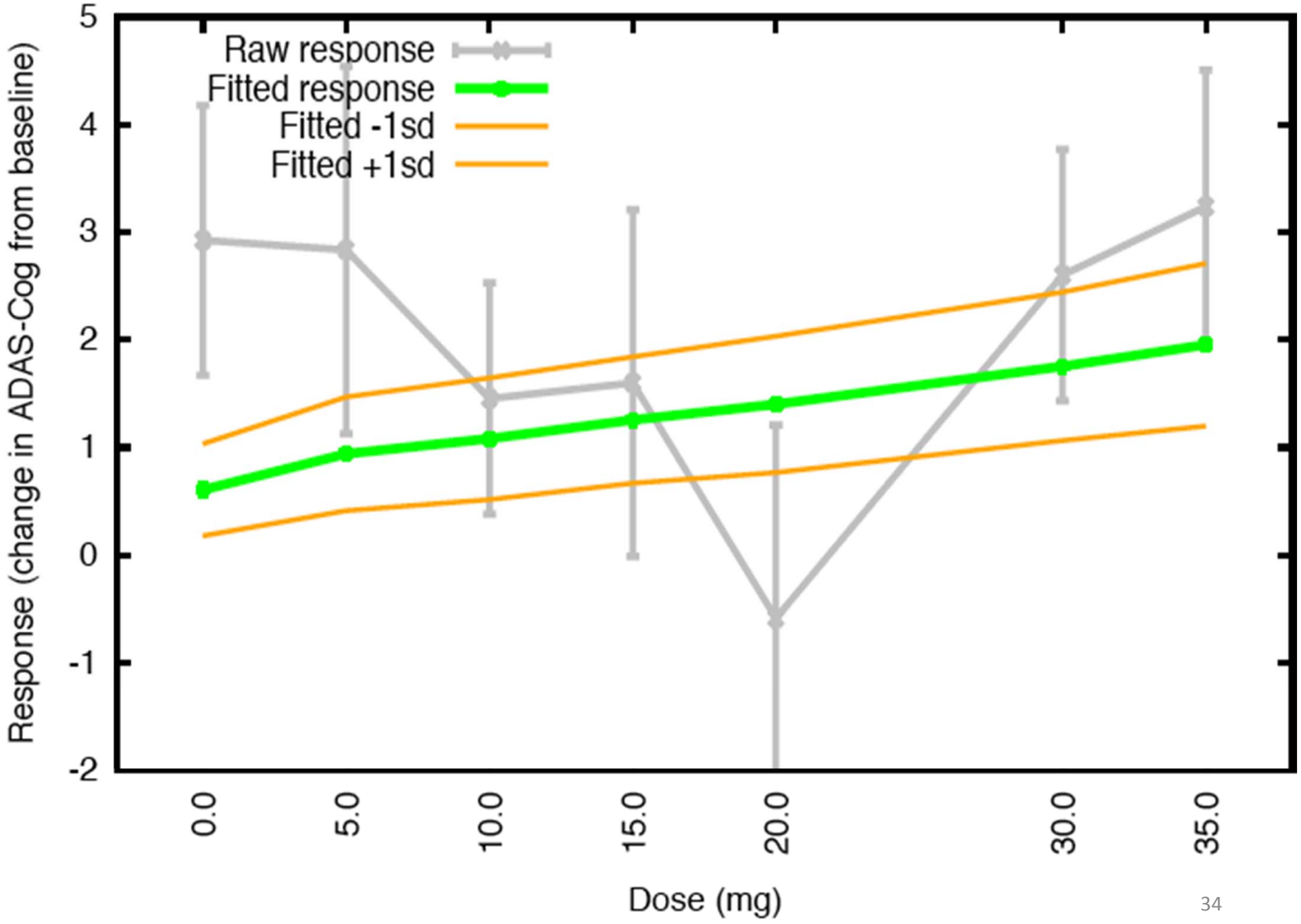
- The primary simulation is based on the best guesses of accrual rate and correlation between early (week 4 and 8) and final time point (week 12) for the ADAS-Cog scores.
- Maximum sample size is 400
- When considering all the simulated D-R curves, very favorable operating characteristics
  - False positive rate less than 5%
  - False negative rate $\leq$ 20%
- Average sample size across D-R relationships is 319

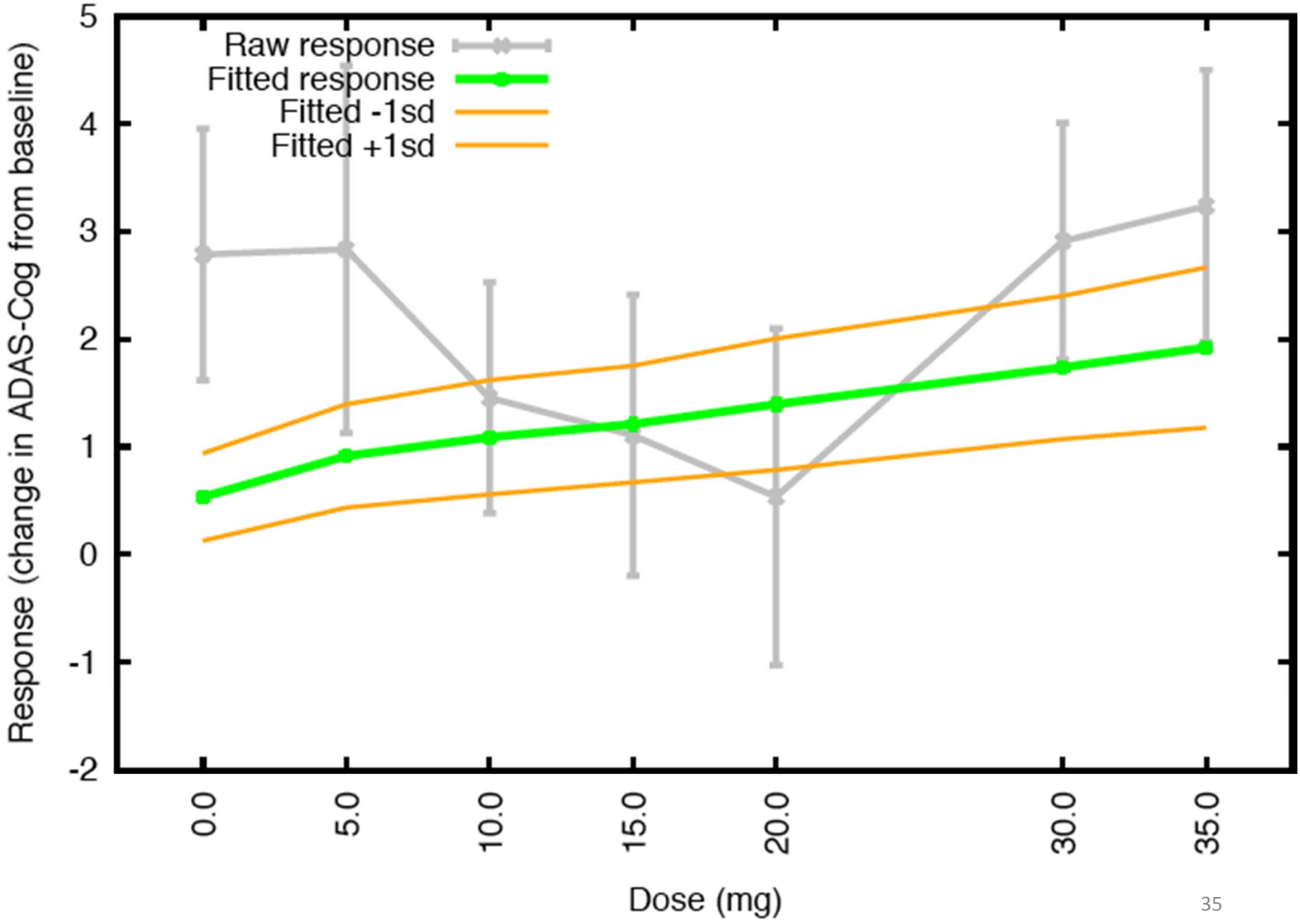
Model estimates of ACAScog (every two weeks)

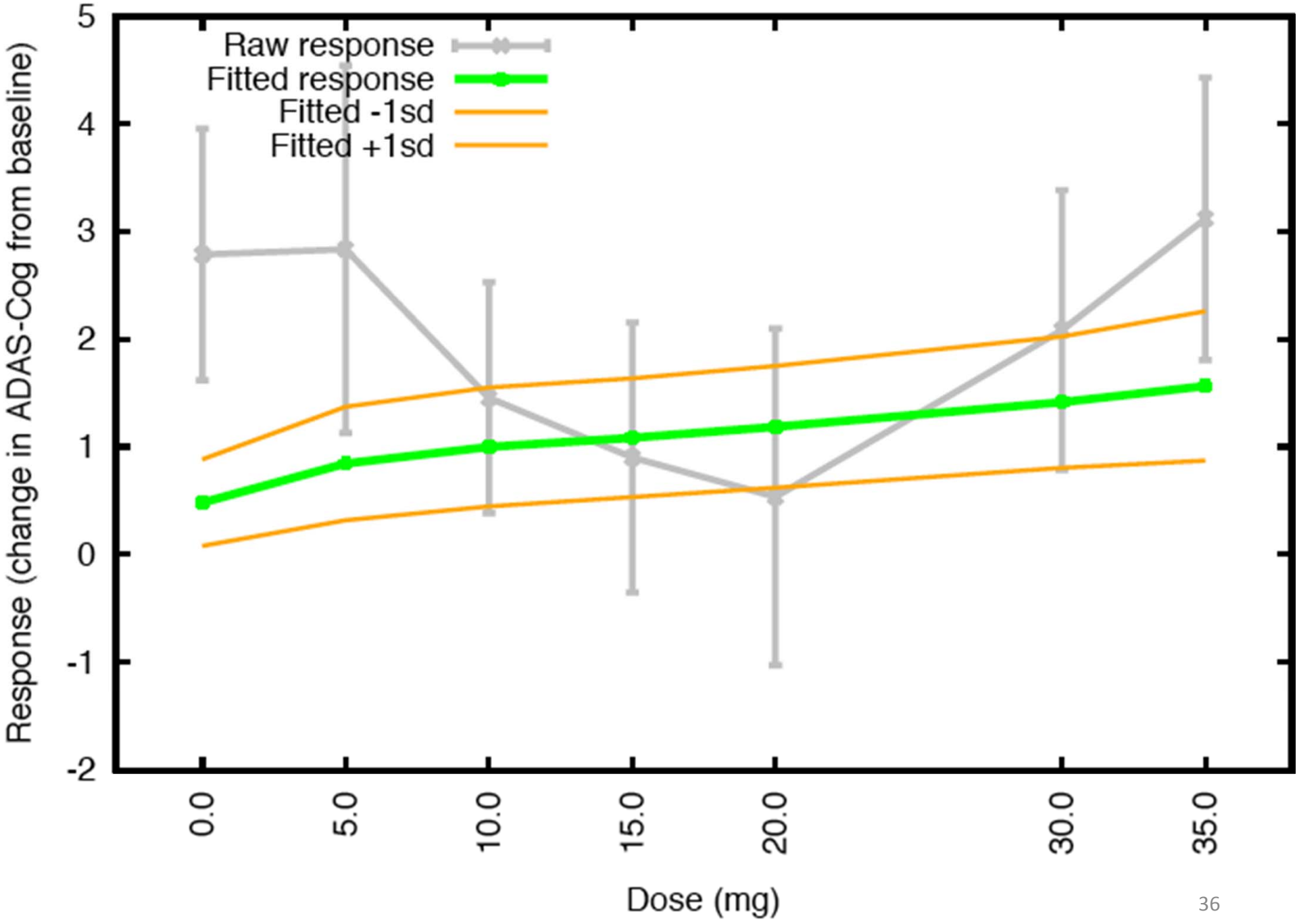

Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

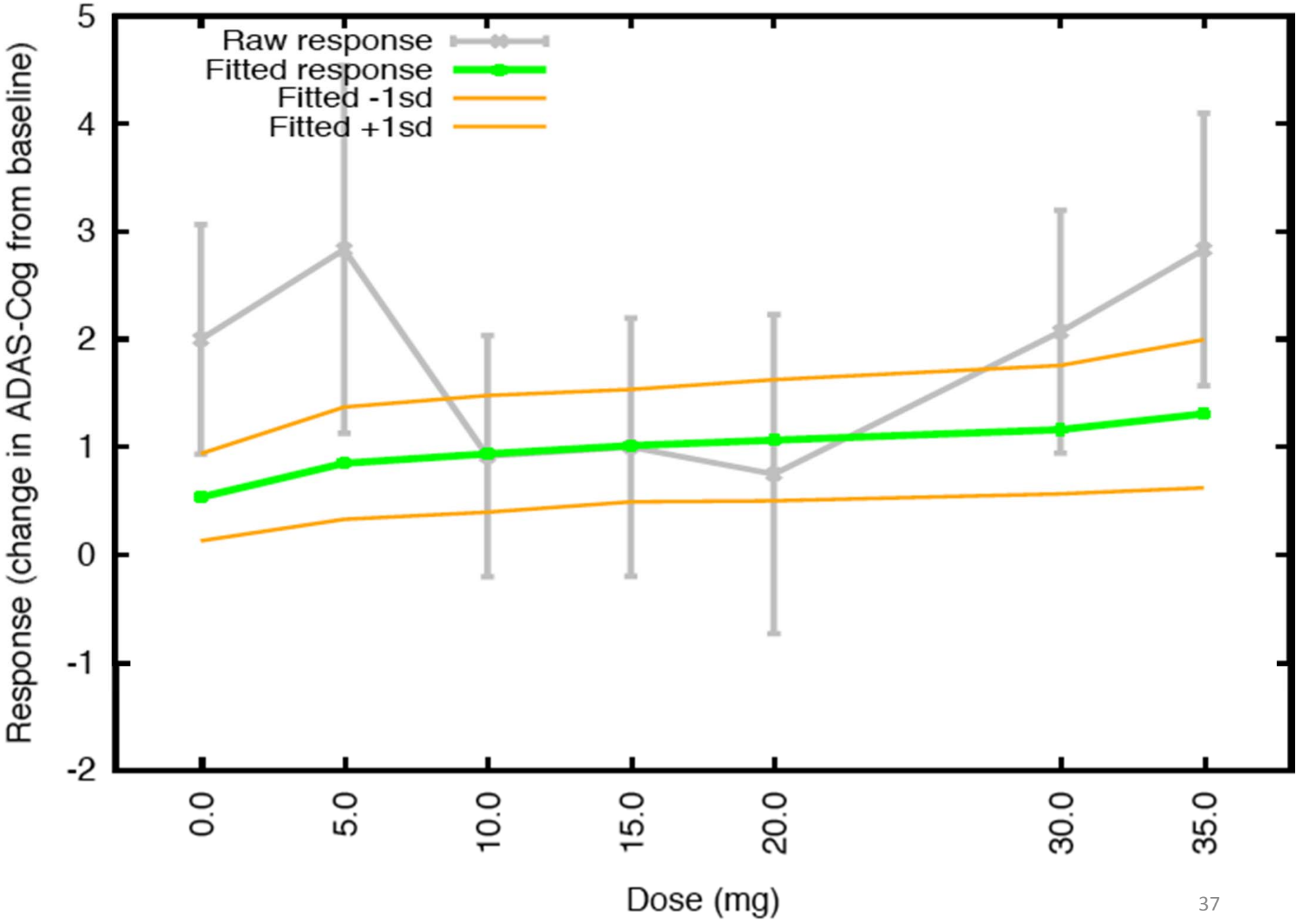

Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

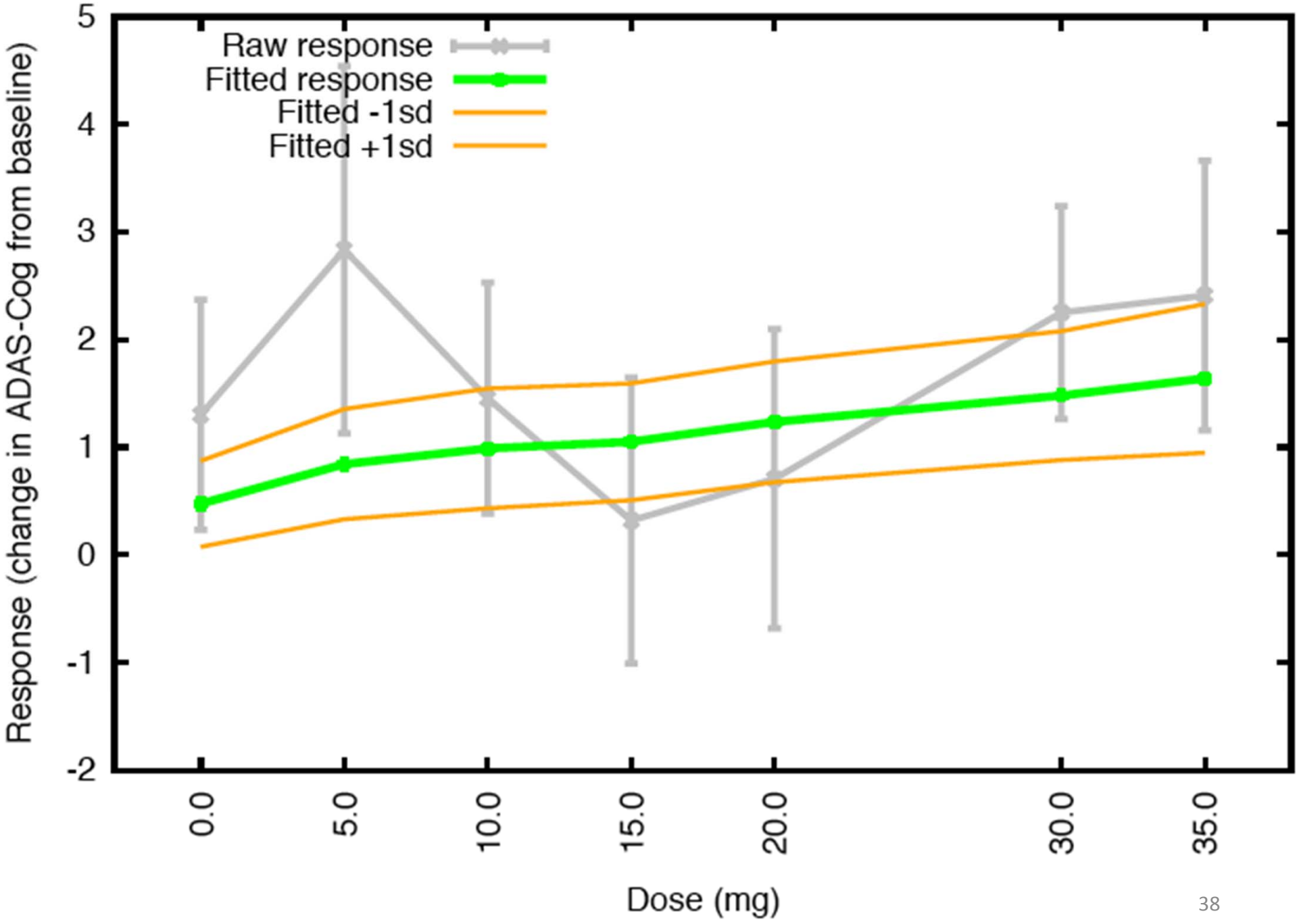

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

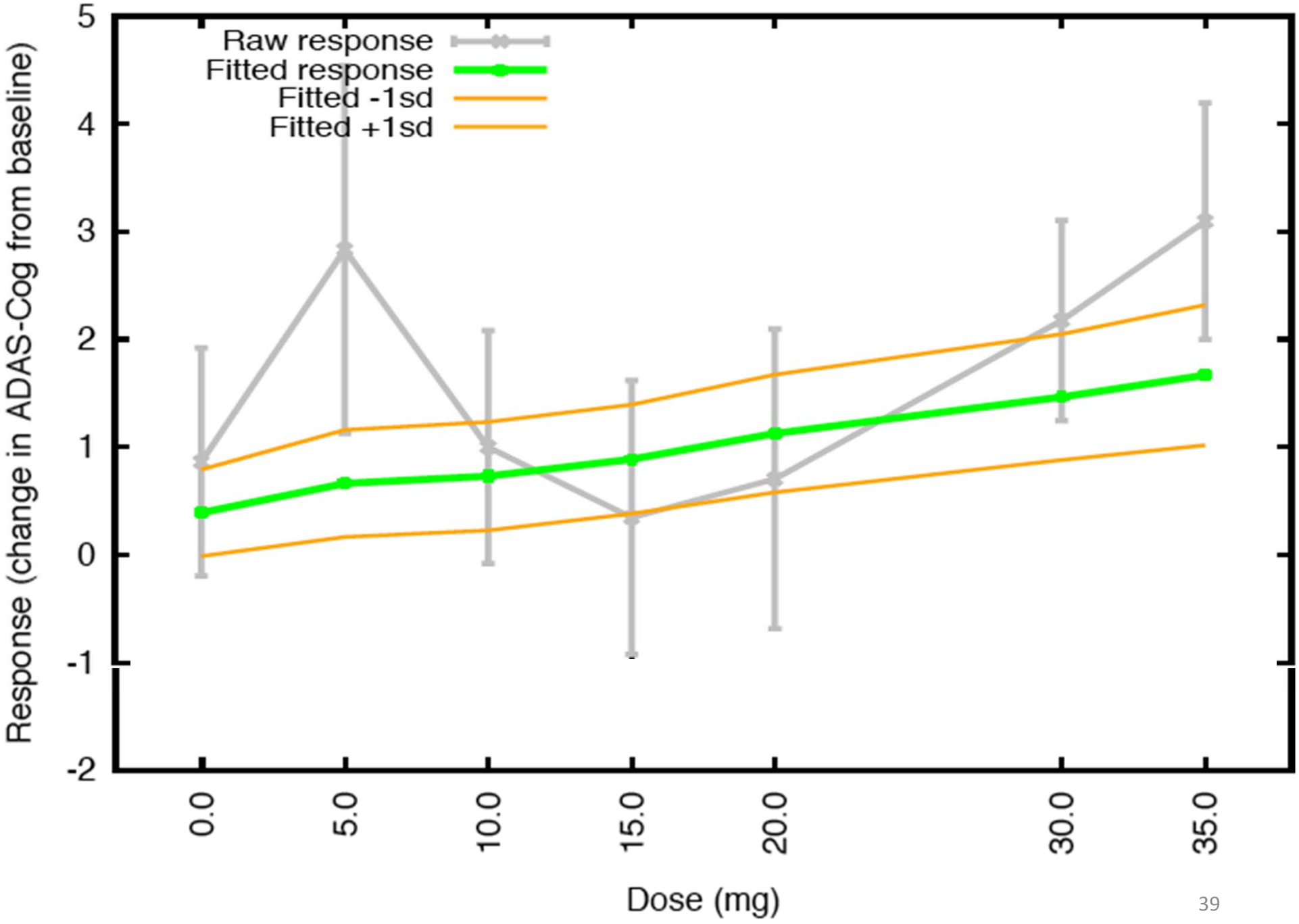

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

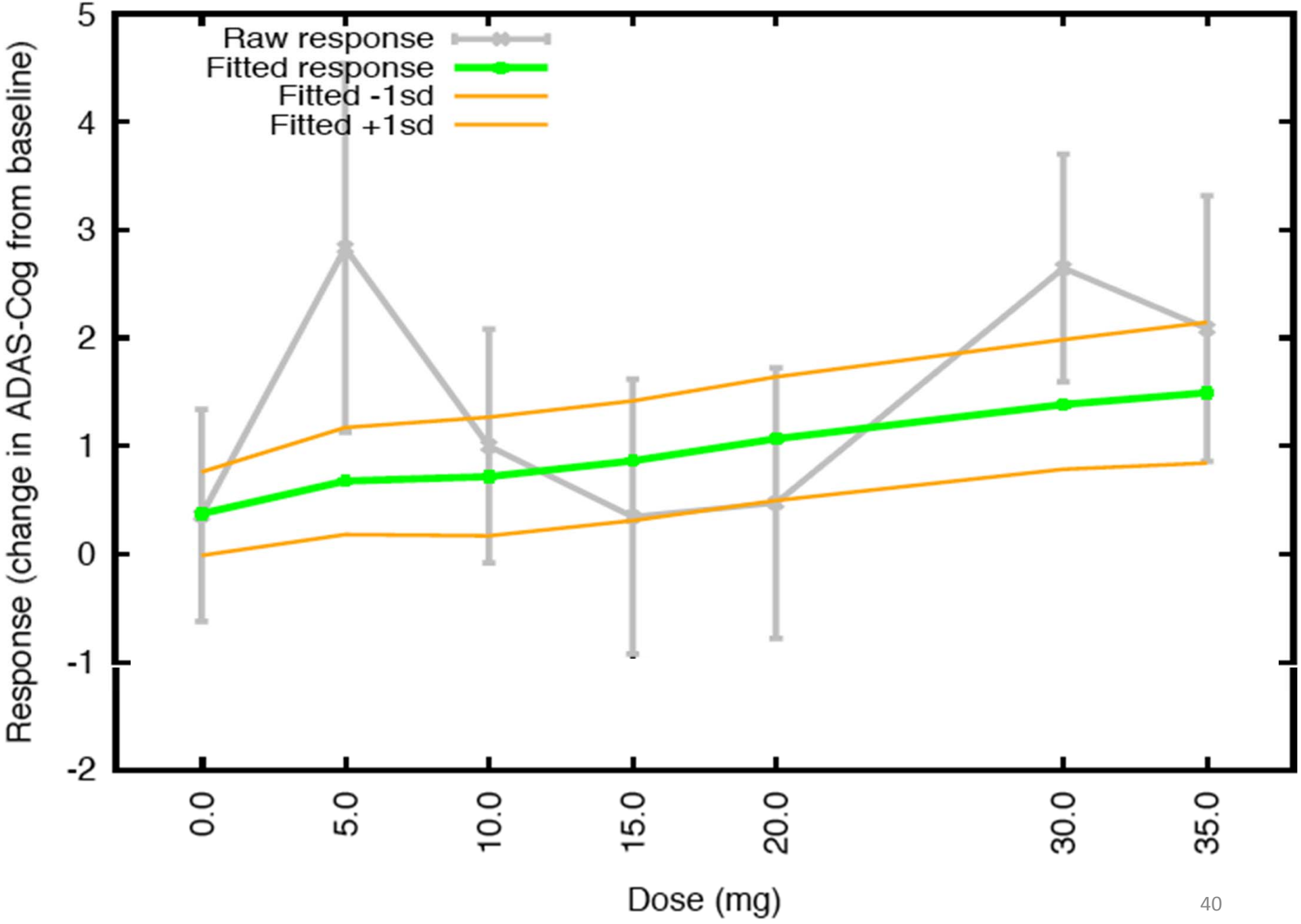

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

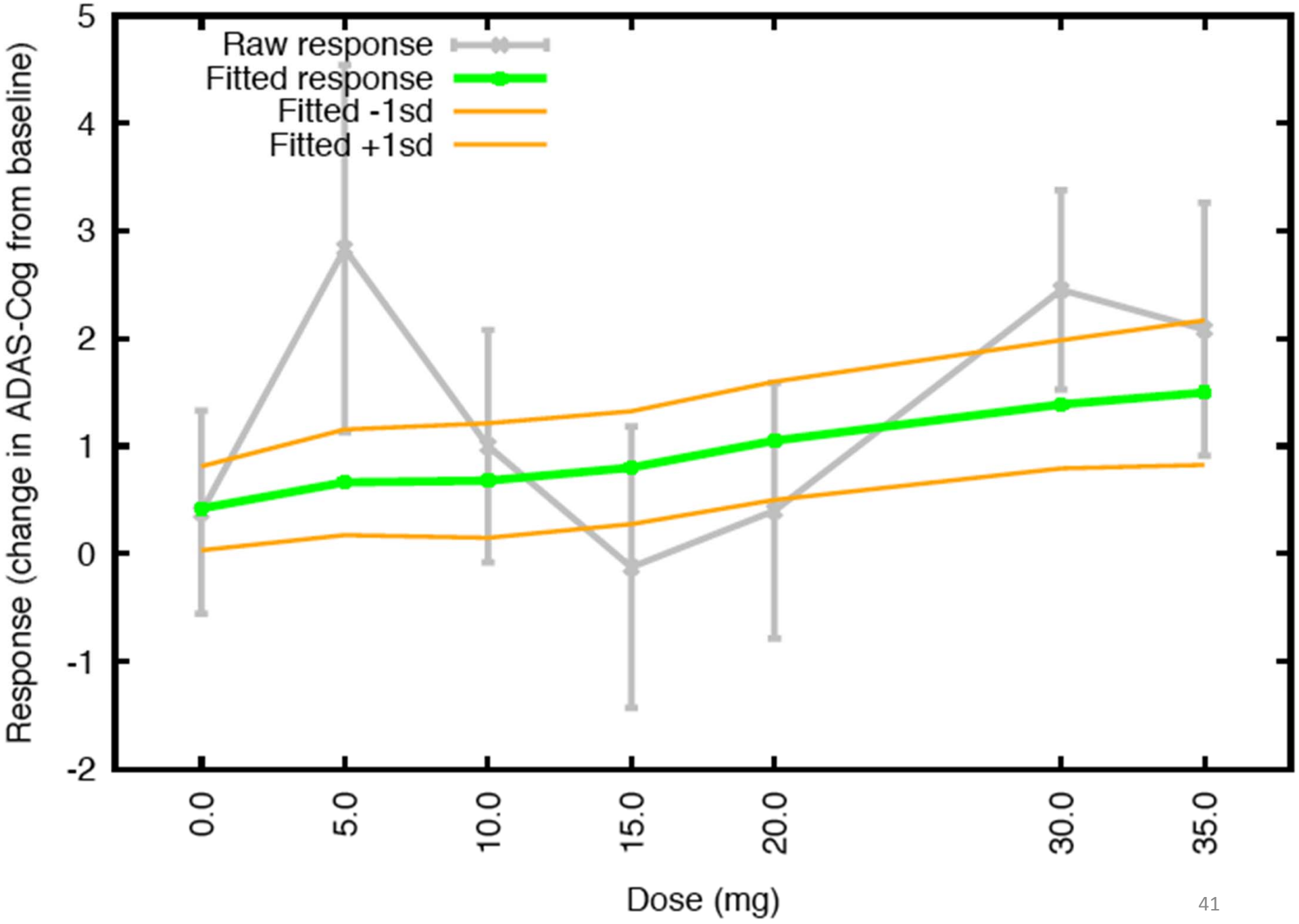

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

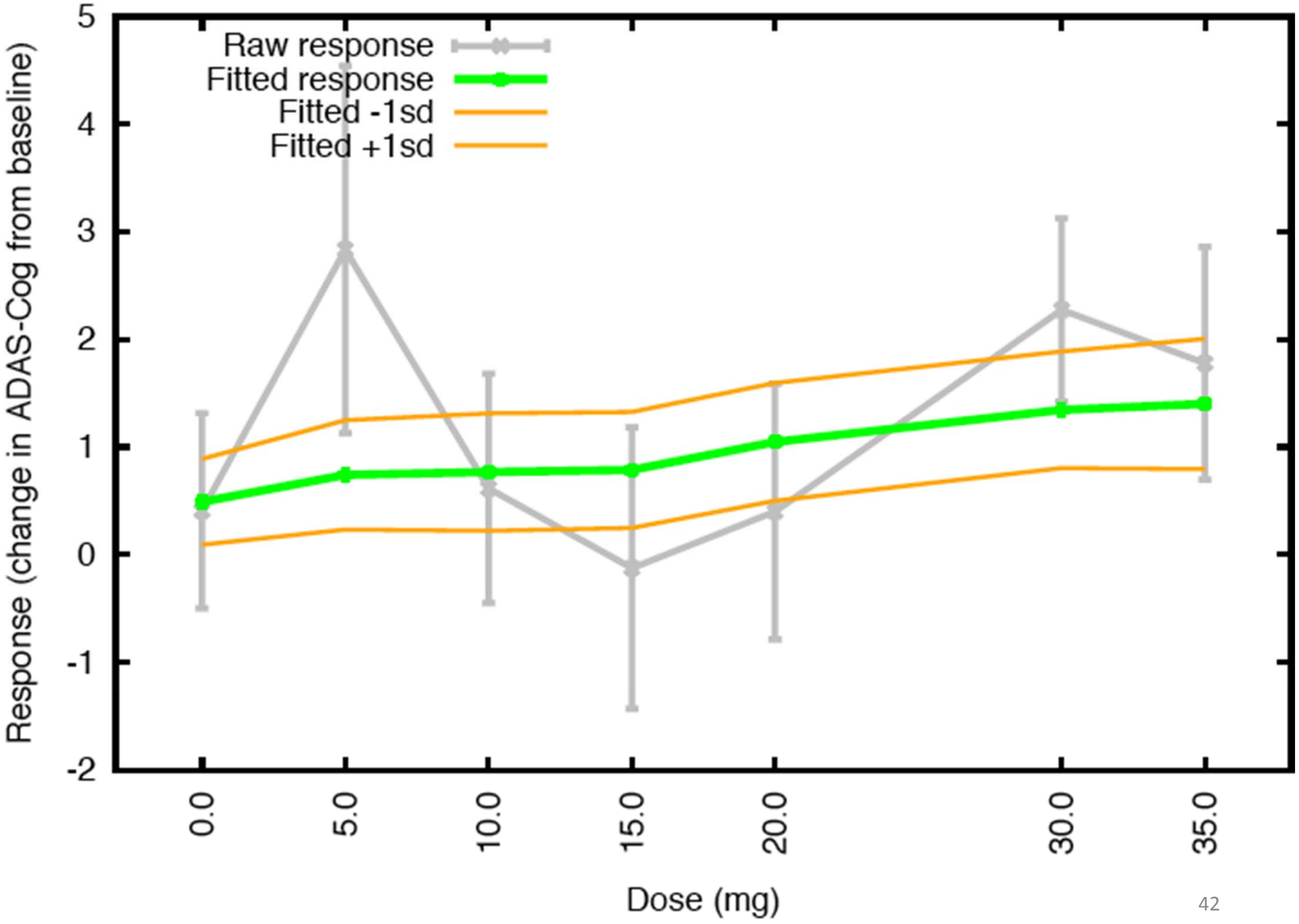

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

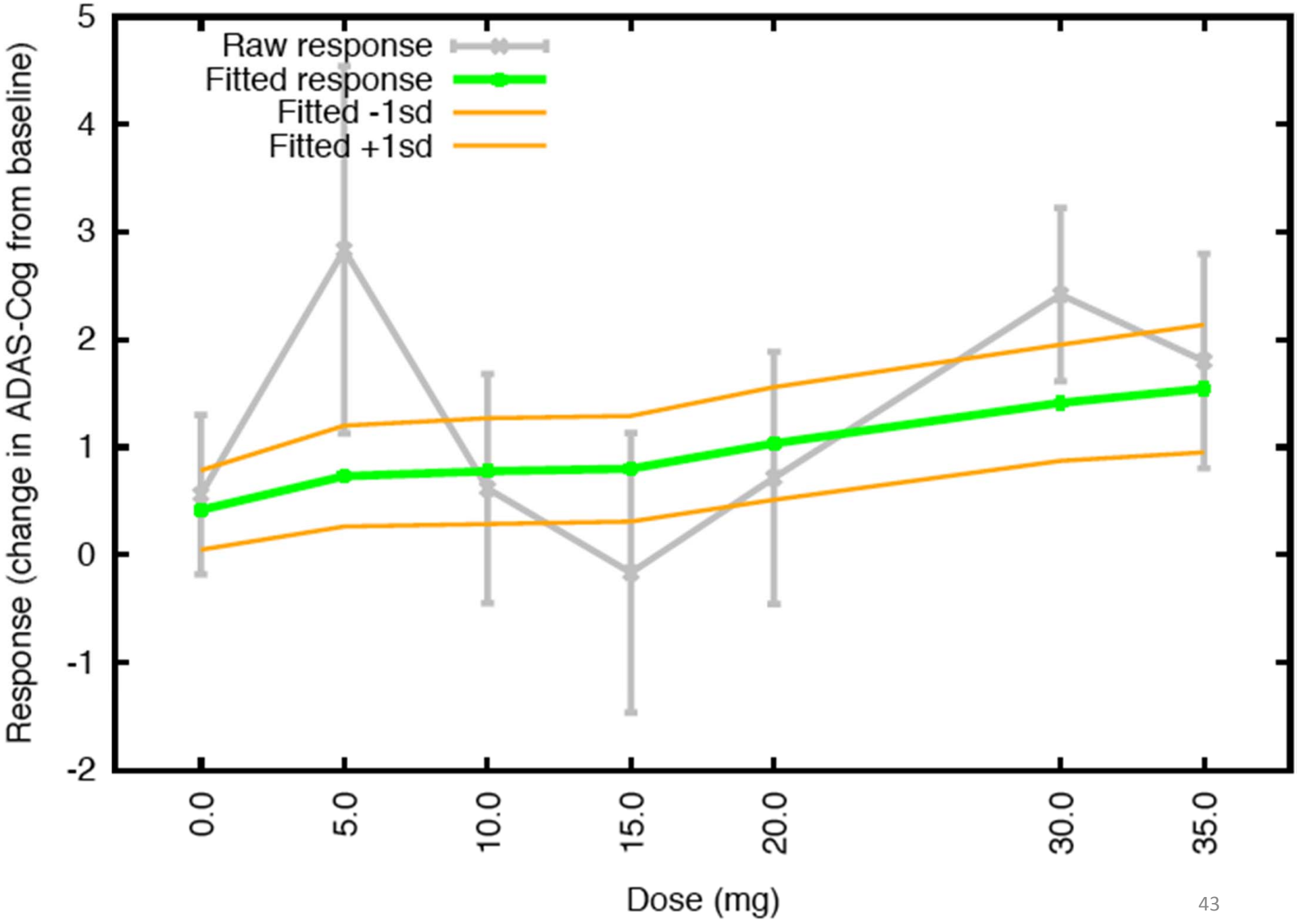

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

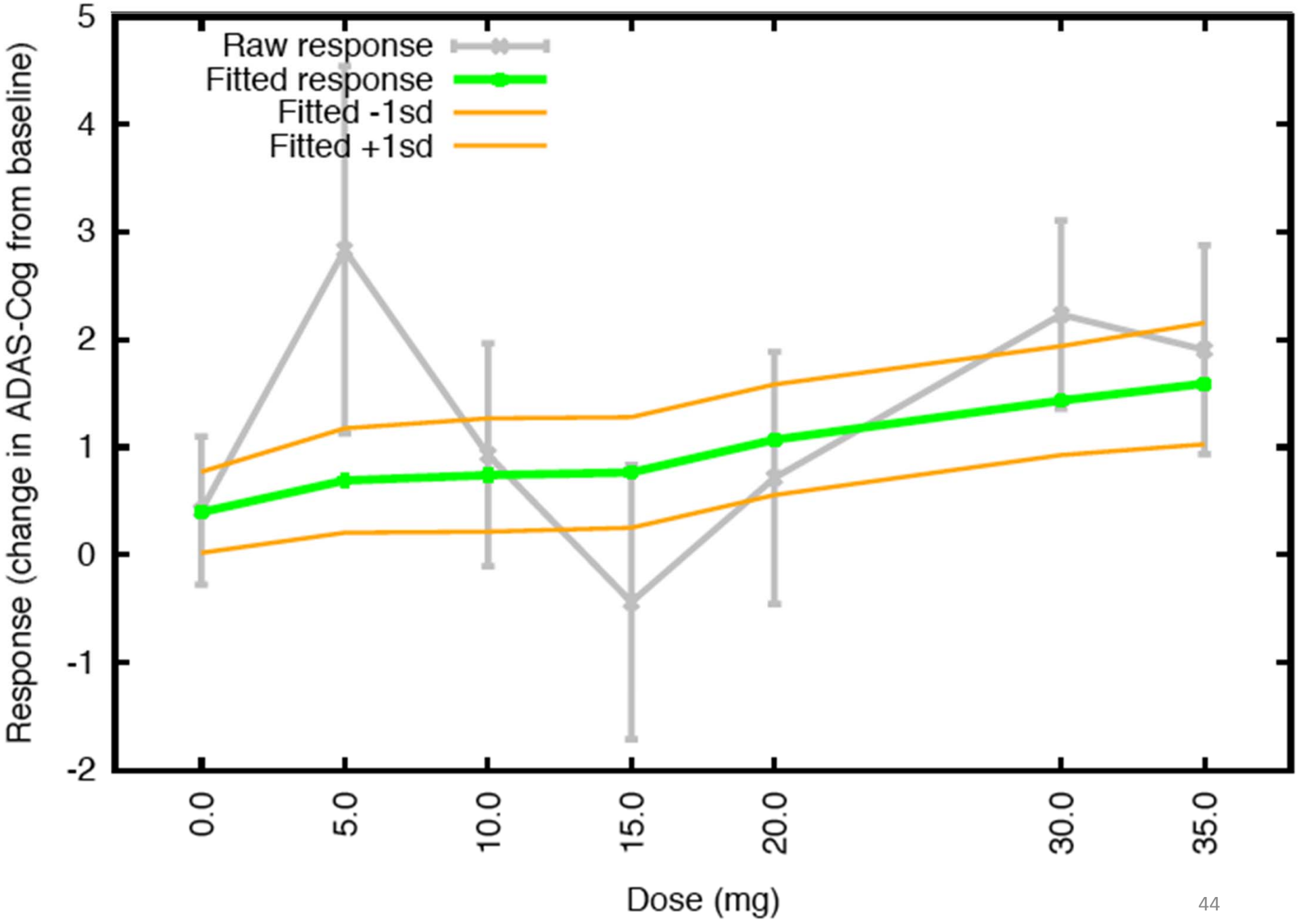

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

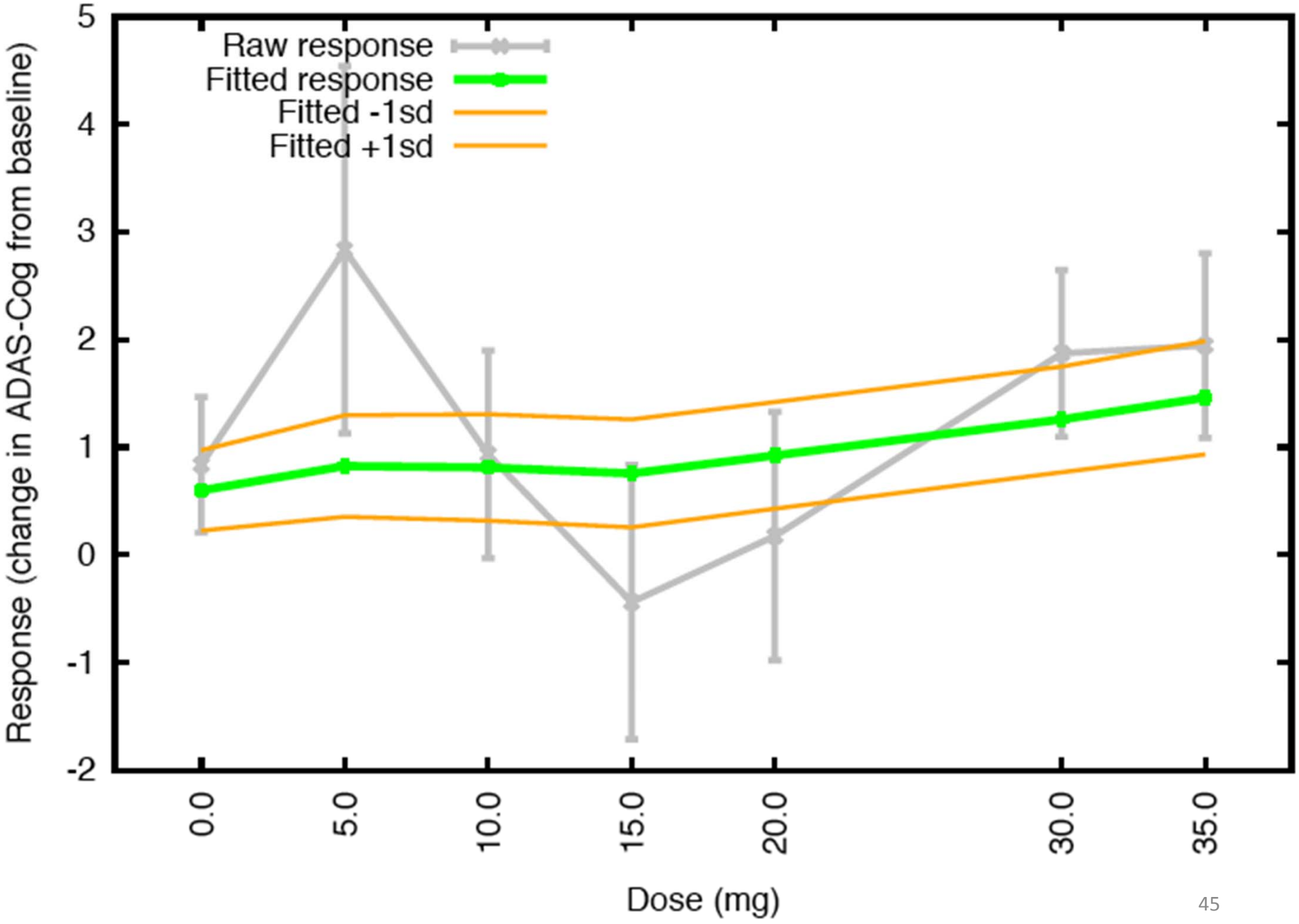

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

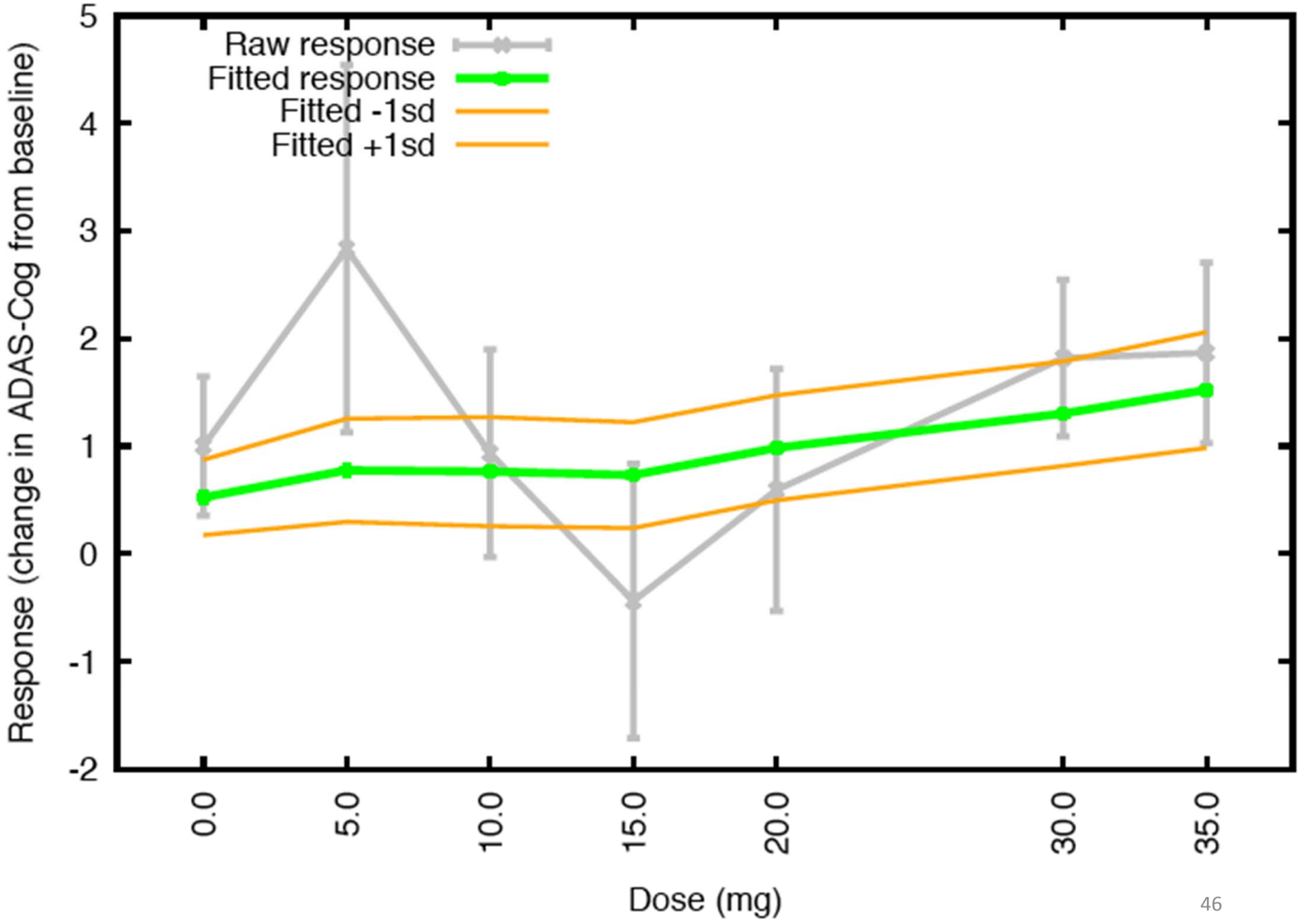

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

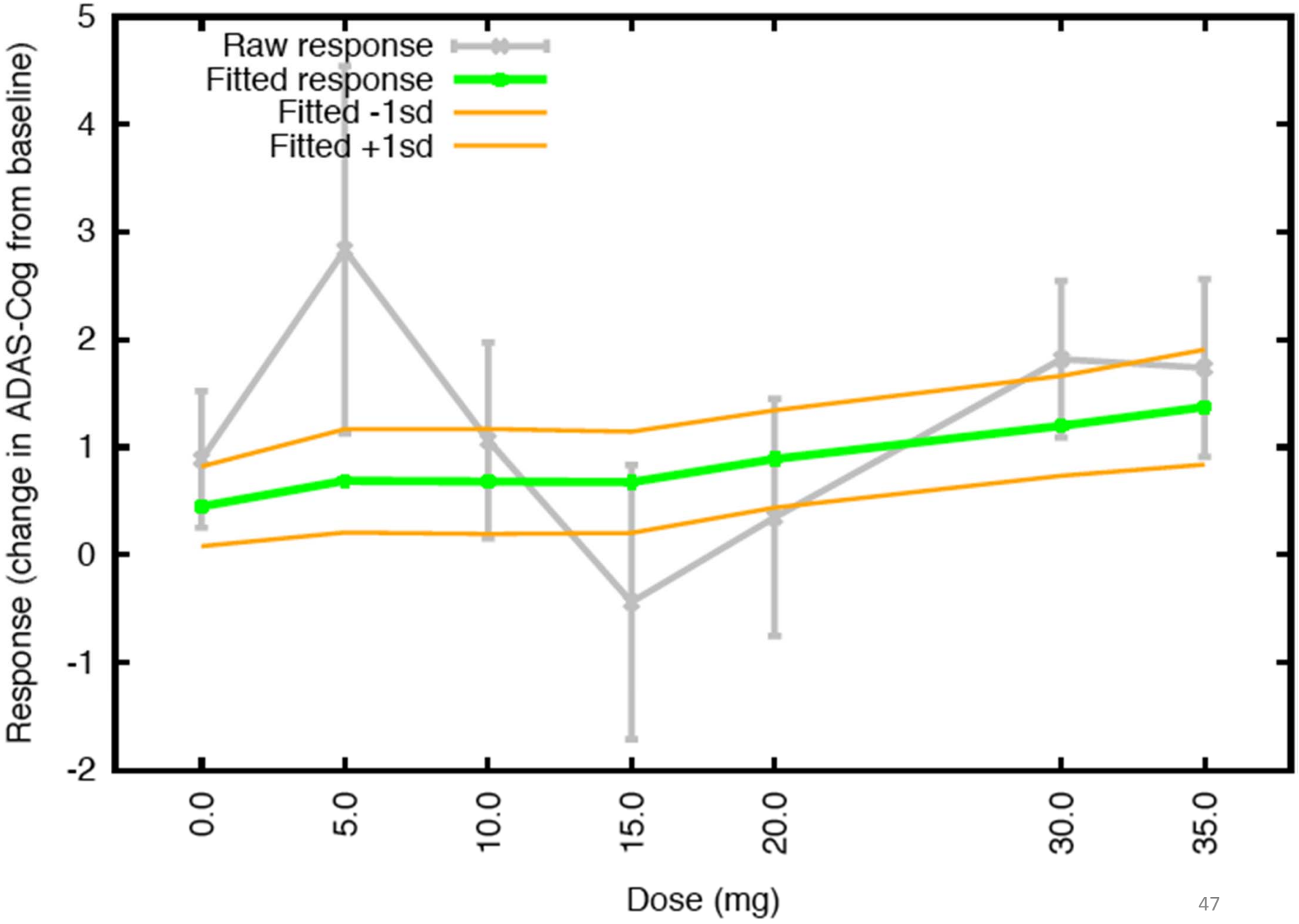

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

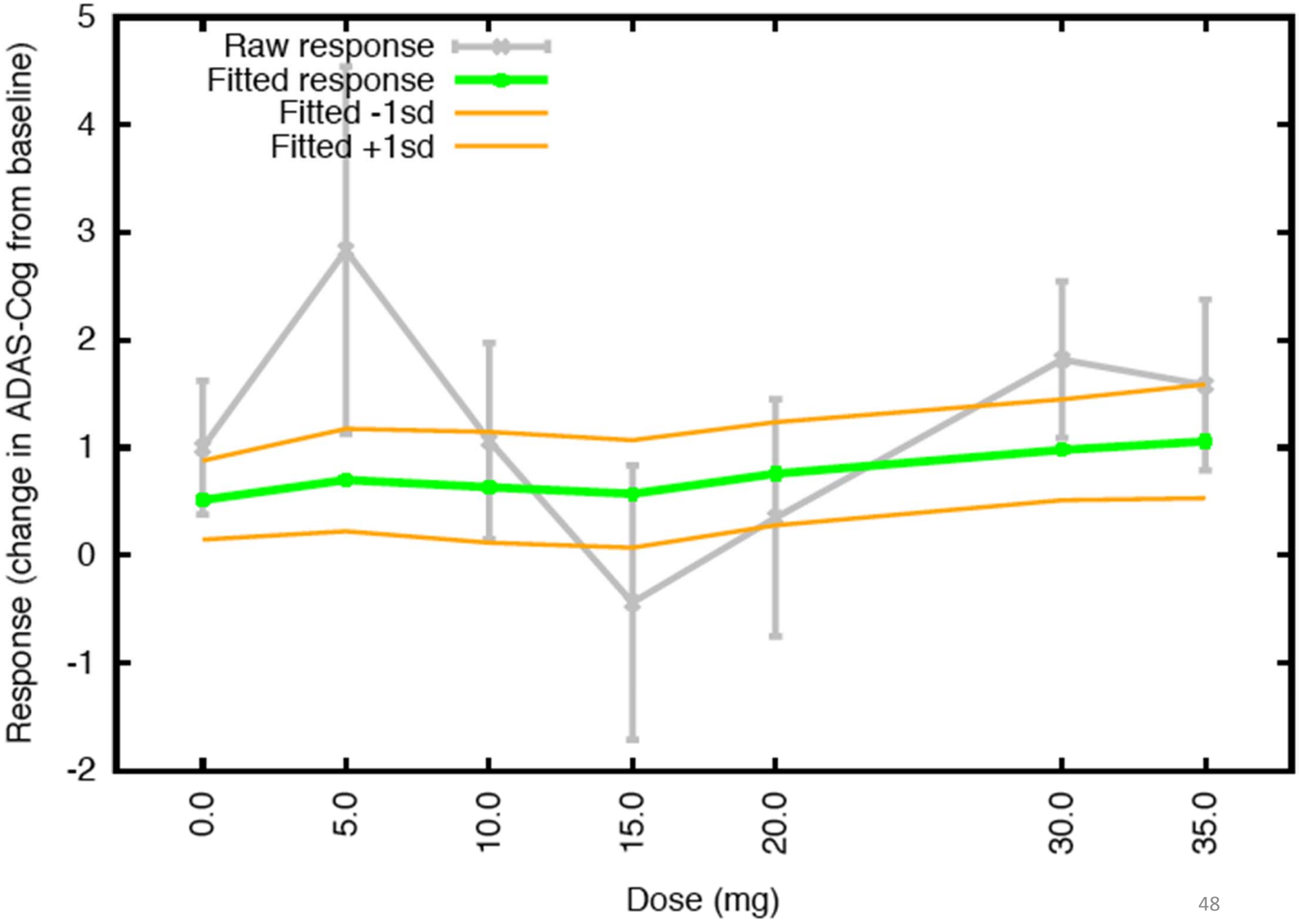

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

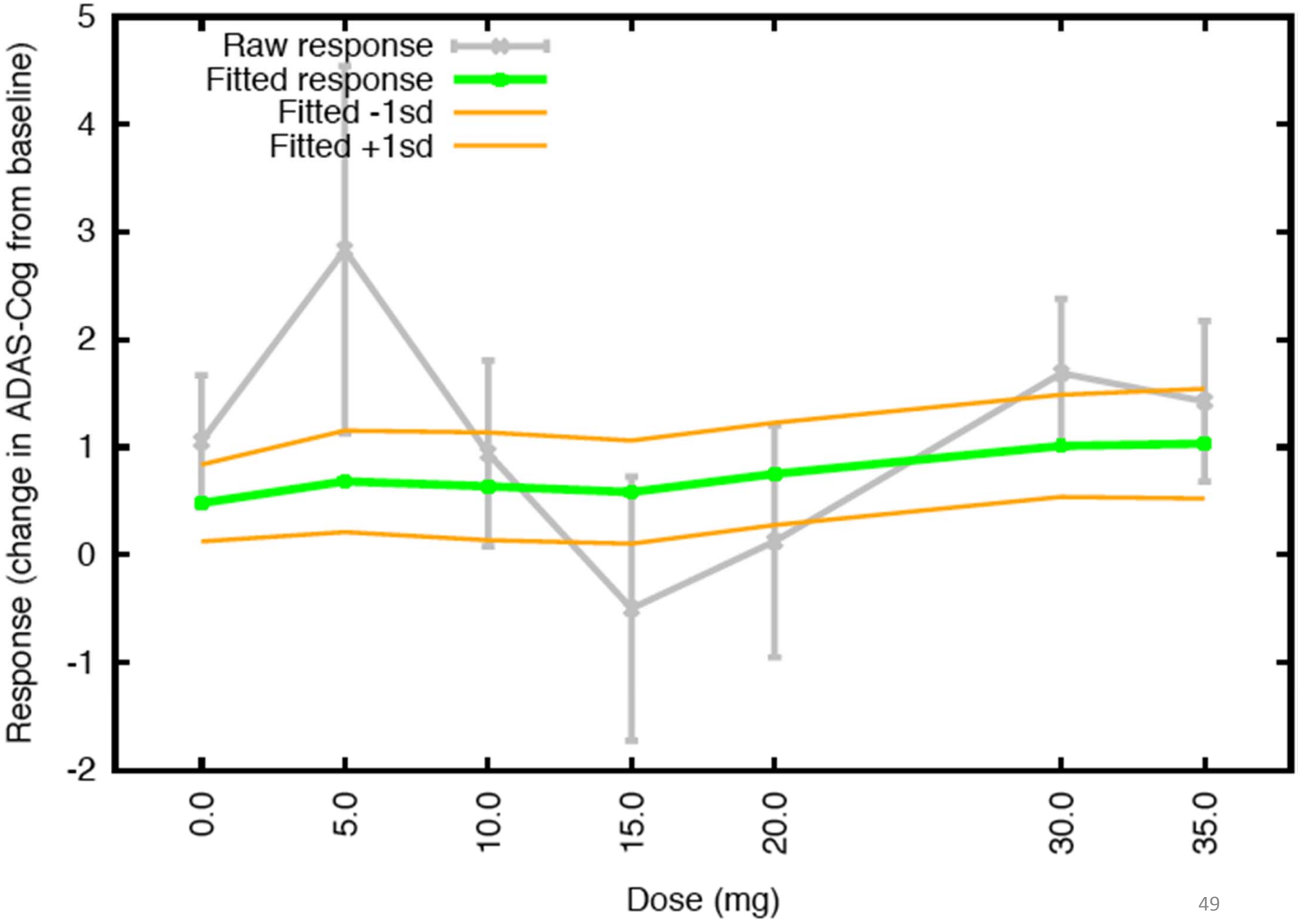

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

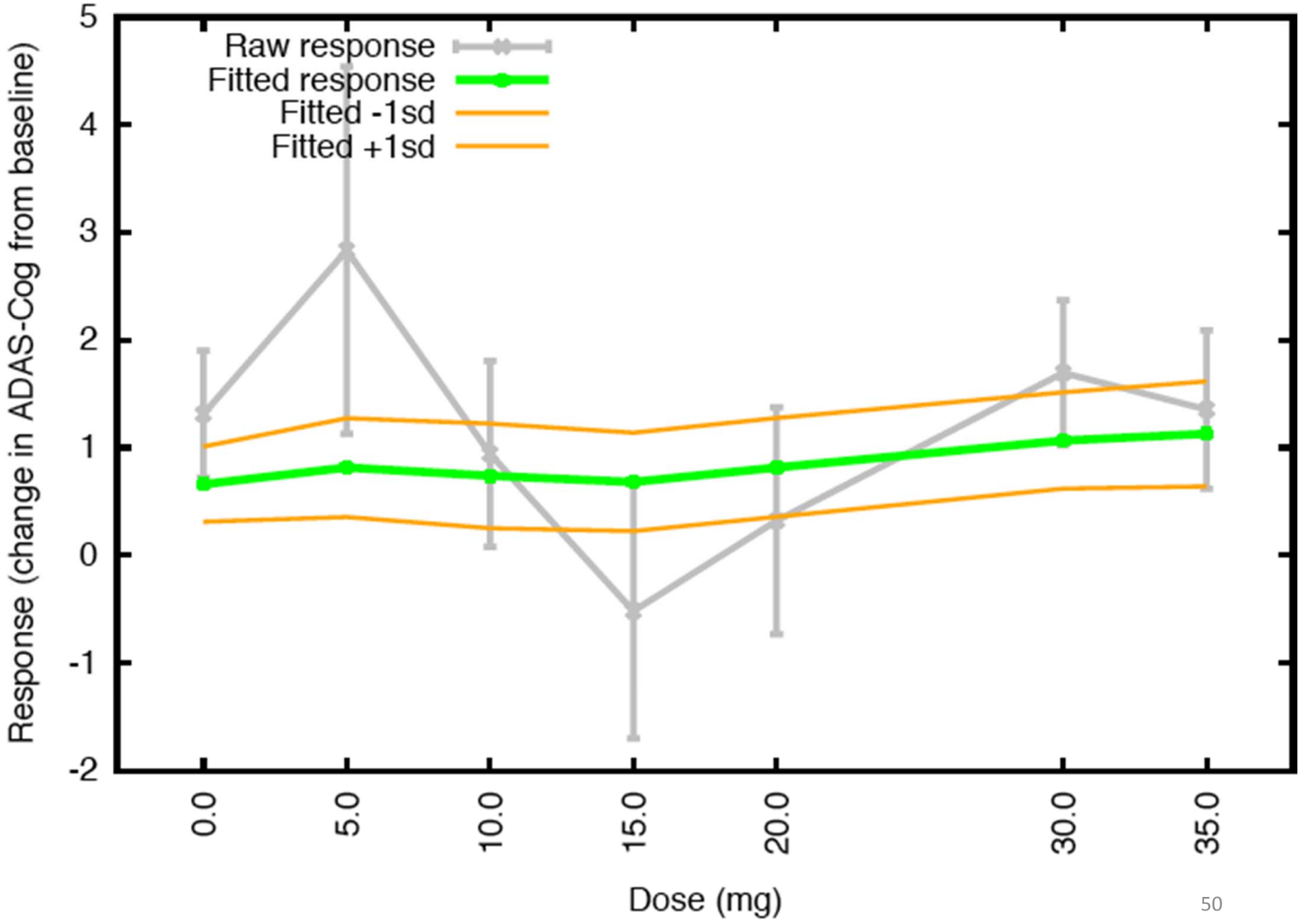

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

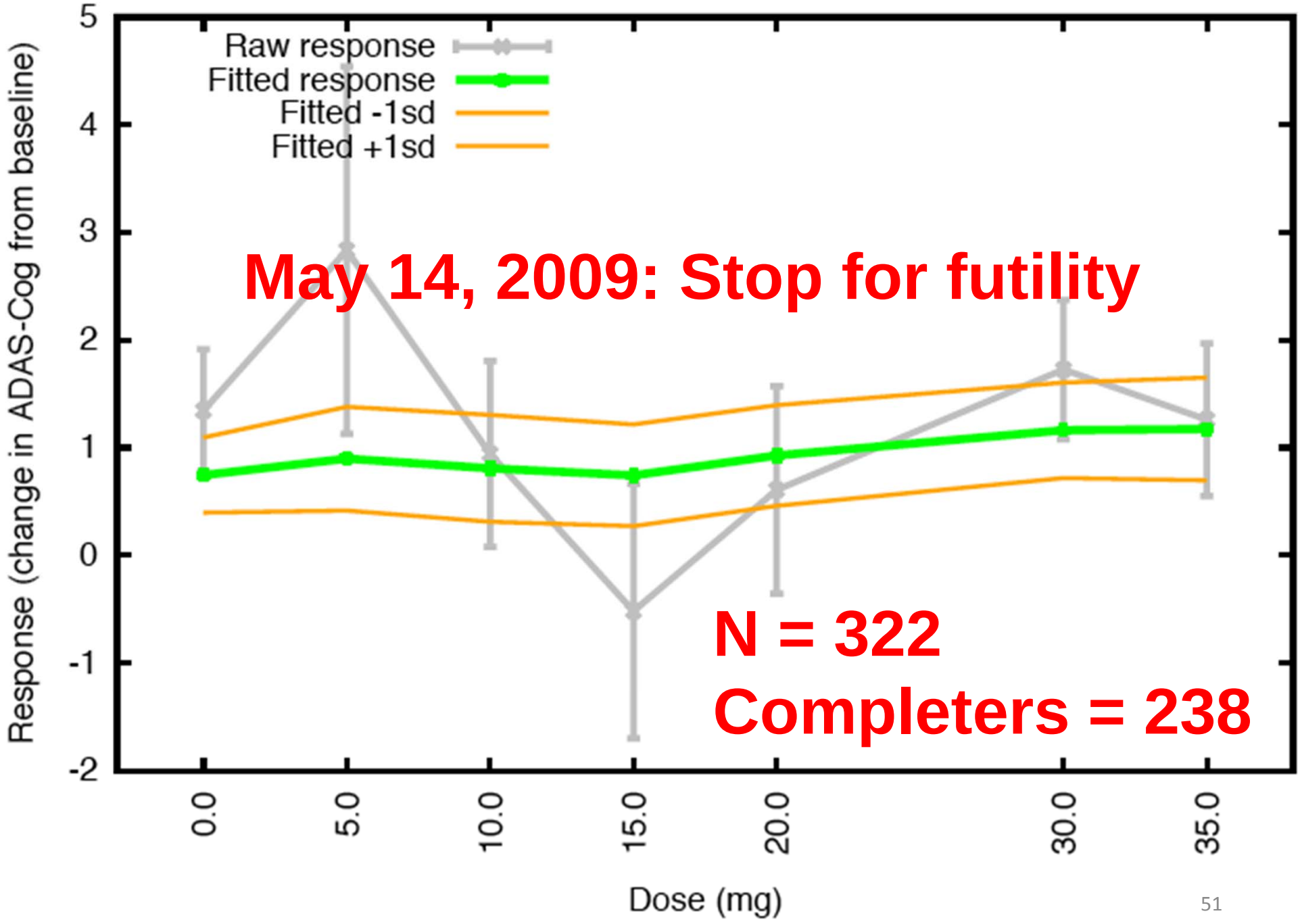

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

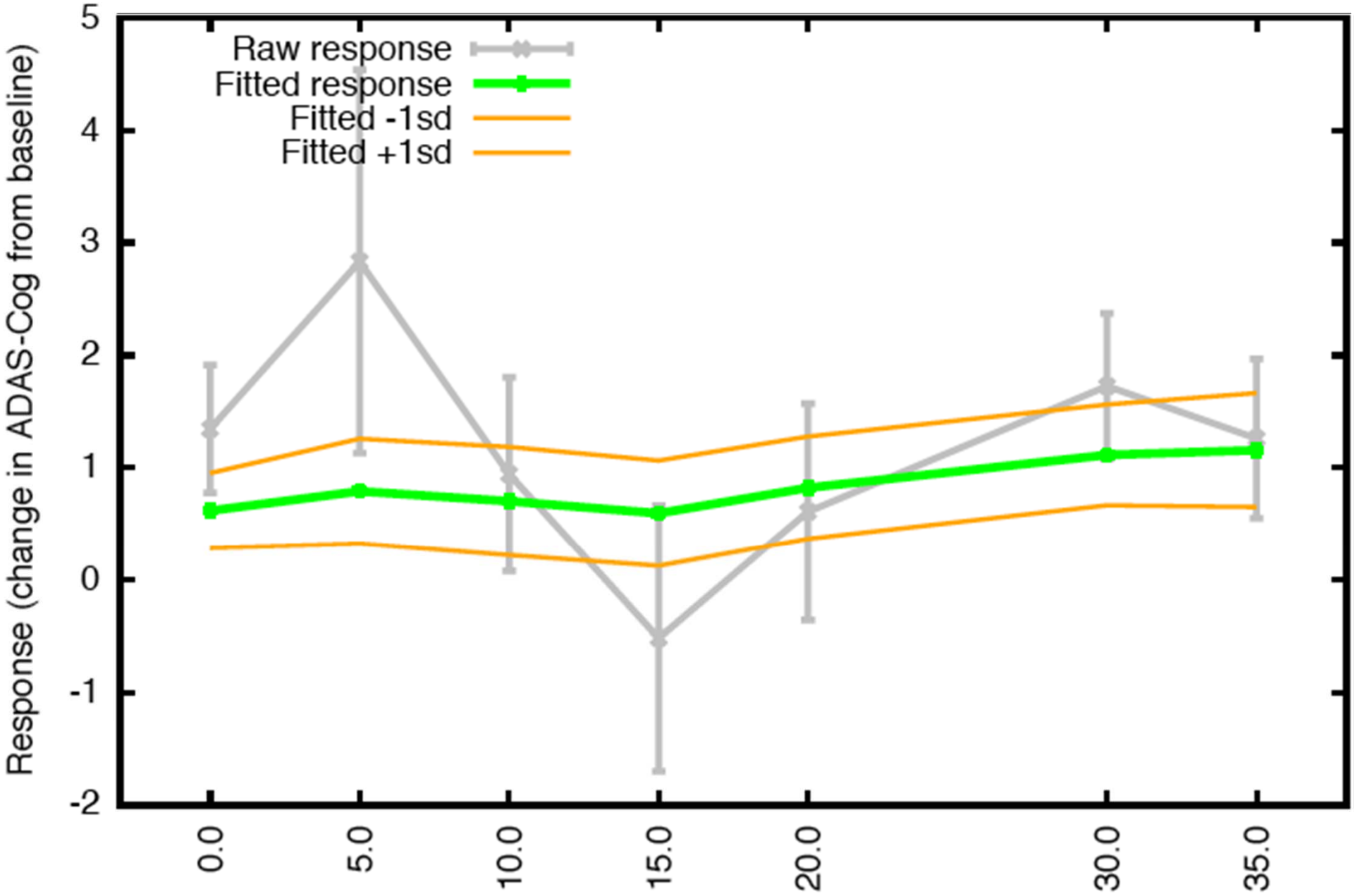

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)


Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

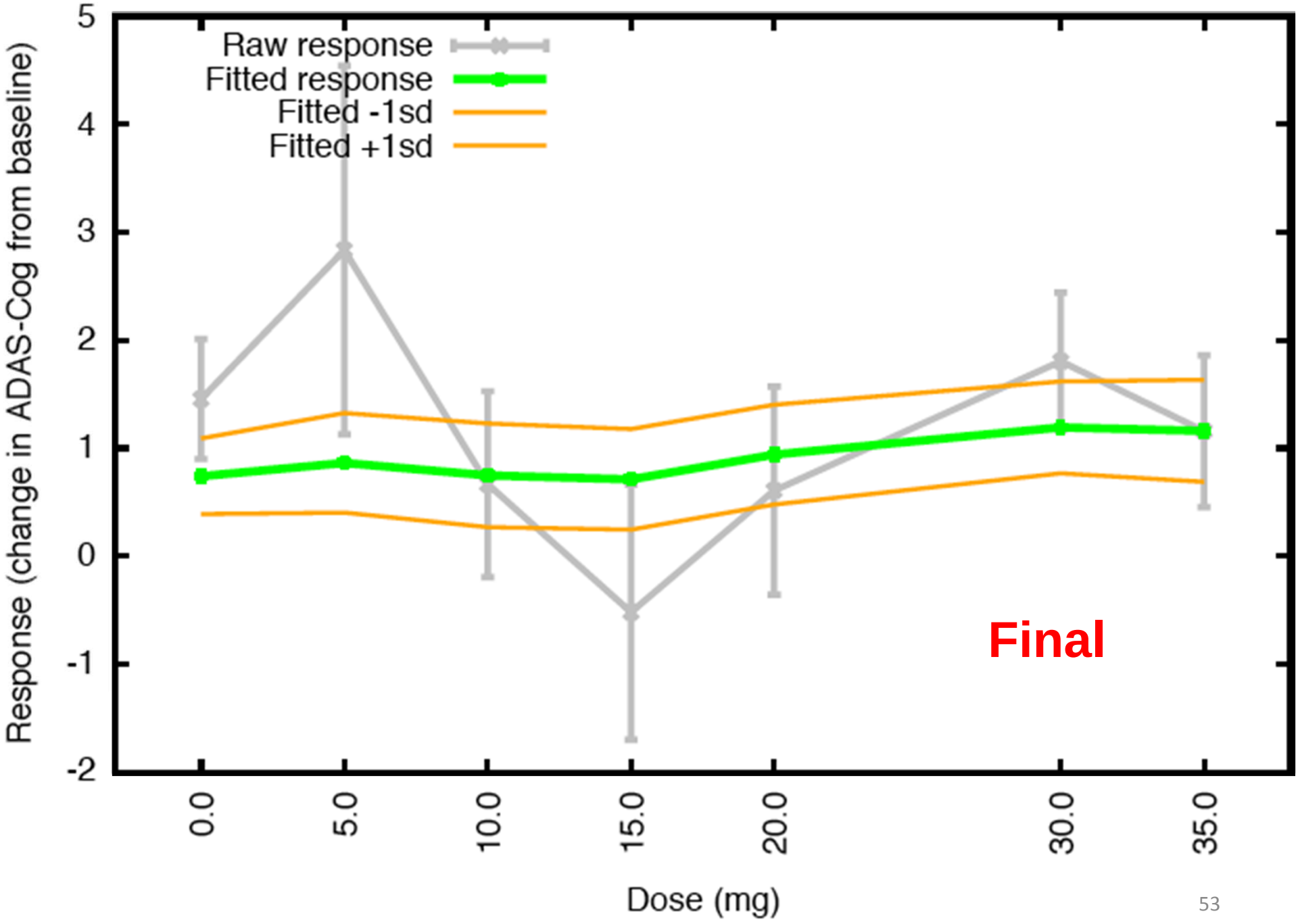

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)


Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

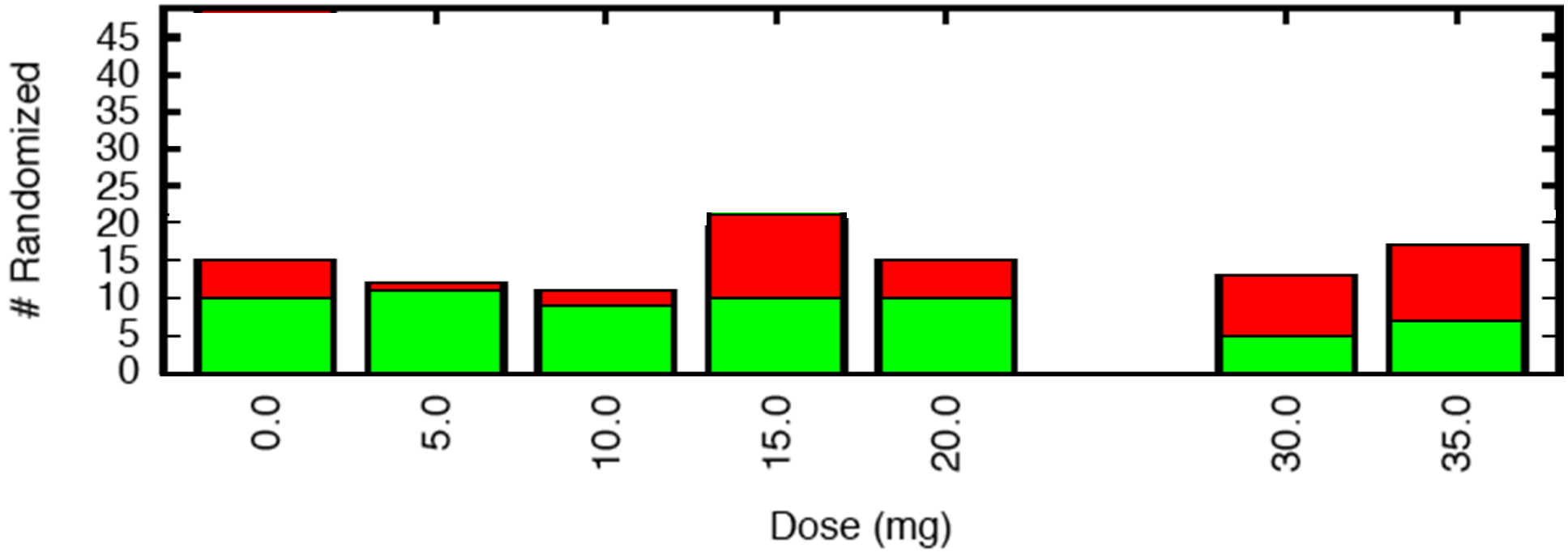

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)
May 14, 2009: Stop for futility
N = 322
Completers = 238

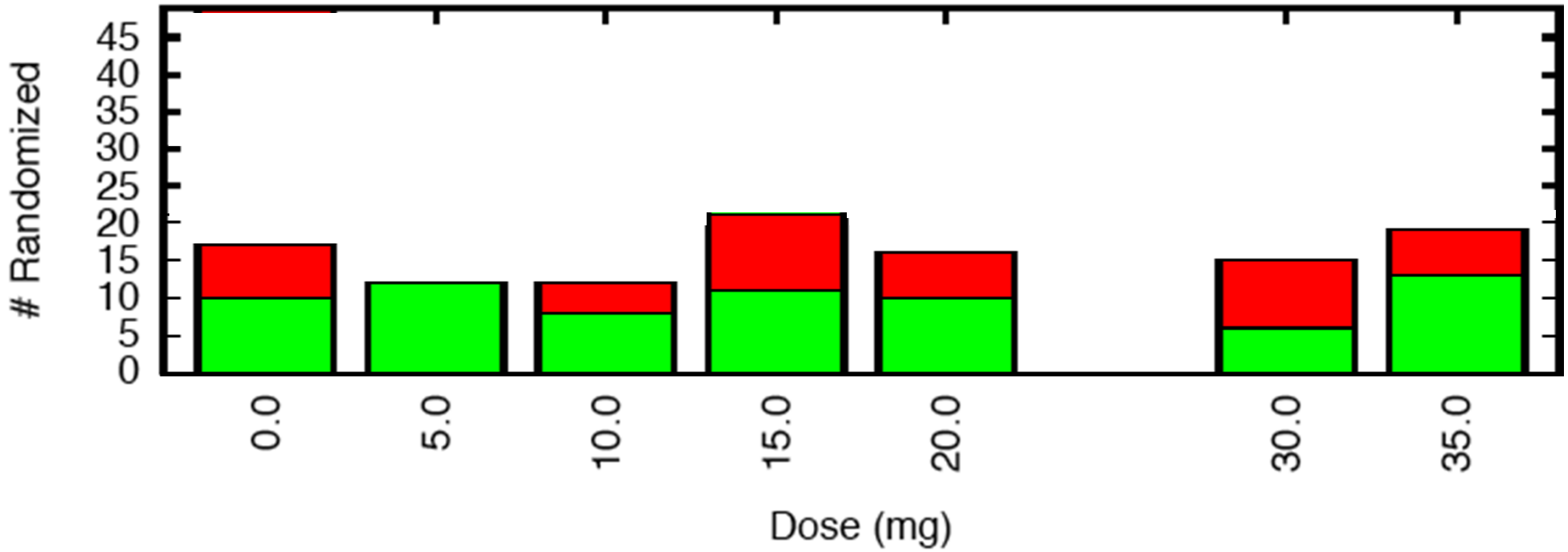


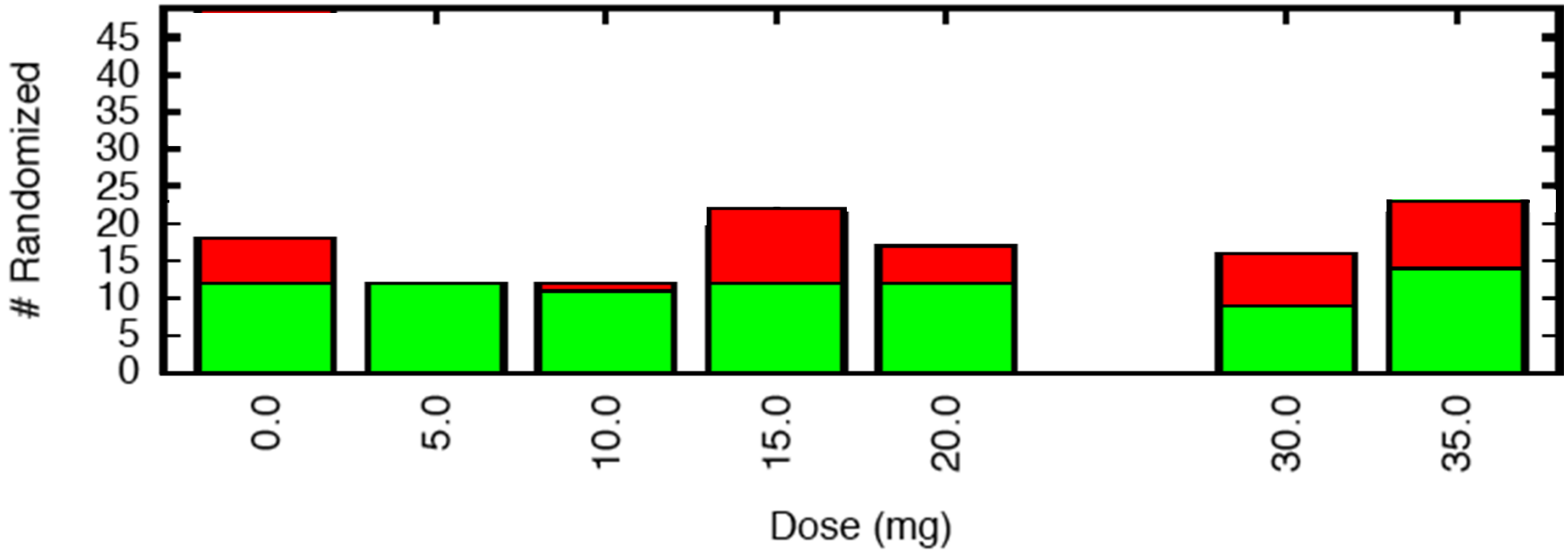
Same as previous, but dropping incomplete patients

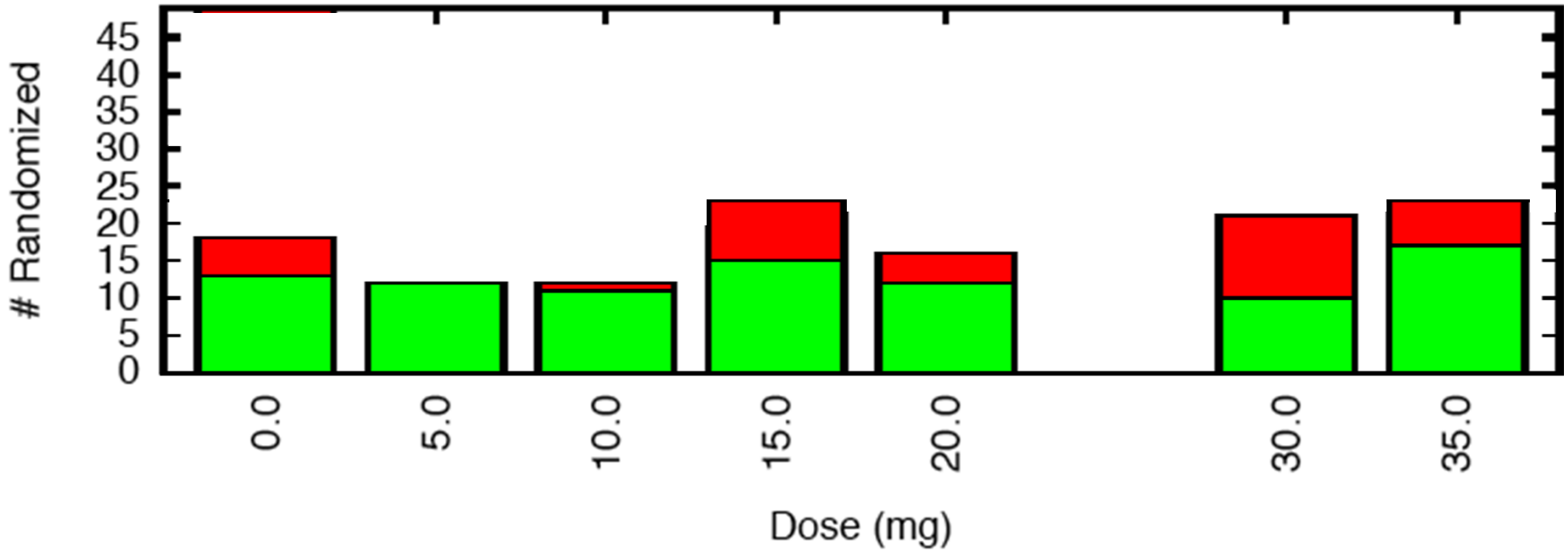

Raw response
Fitted response
Fitted -1sd
Fitted +1sd
Response (change in ADAS-Cog from baseline)
5
4
3
2
1
0
-1
-2
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)
Final

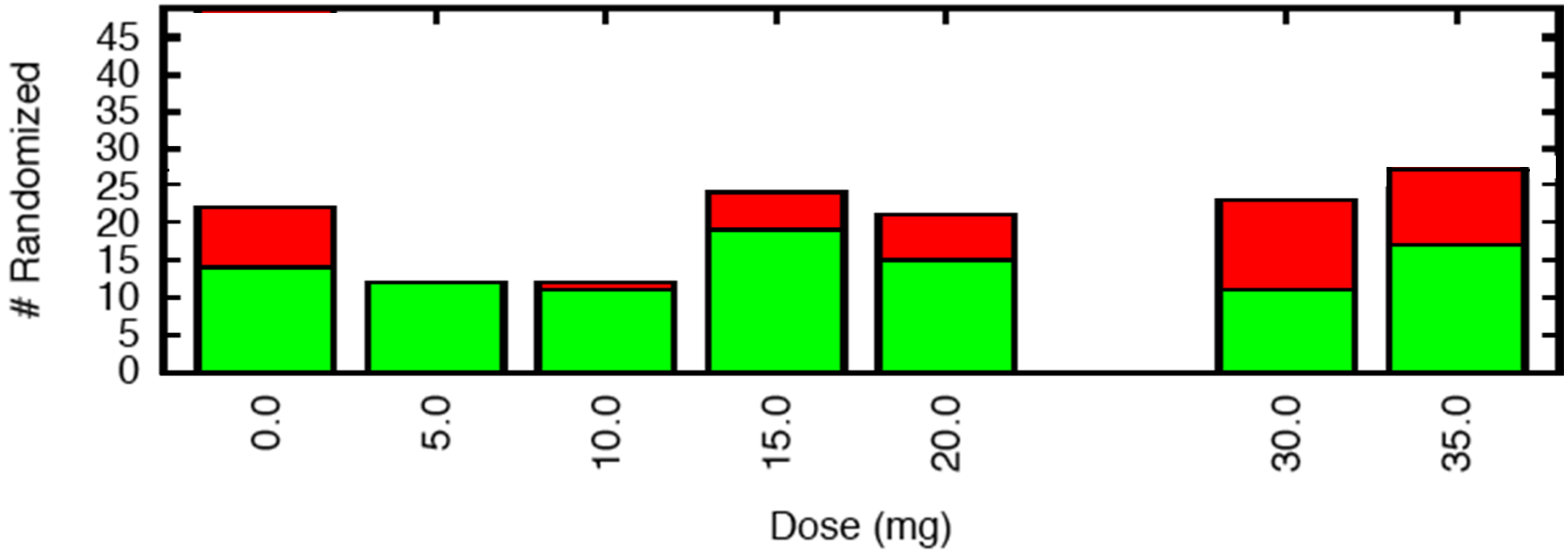
Sample sizes

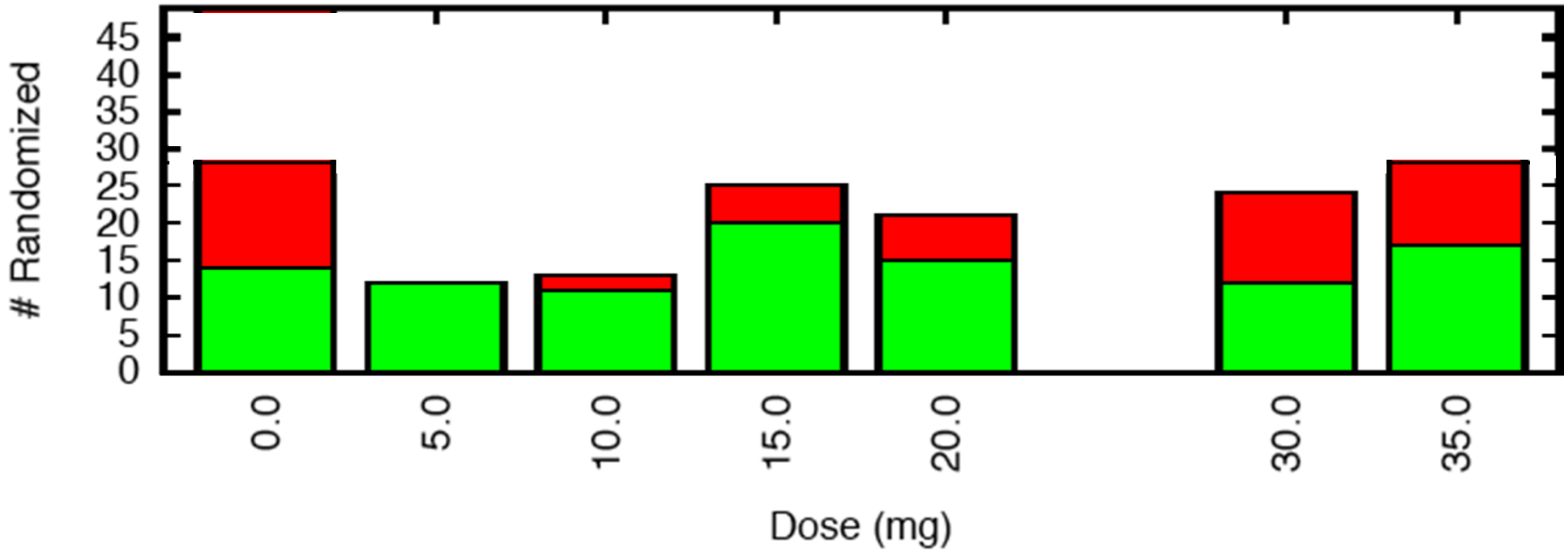

Randomized
45
40
35
30
25
20
15
10
5
0
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

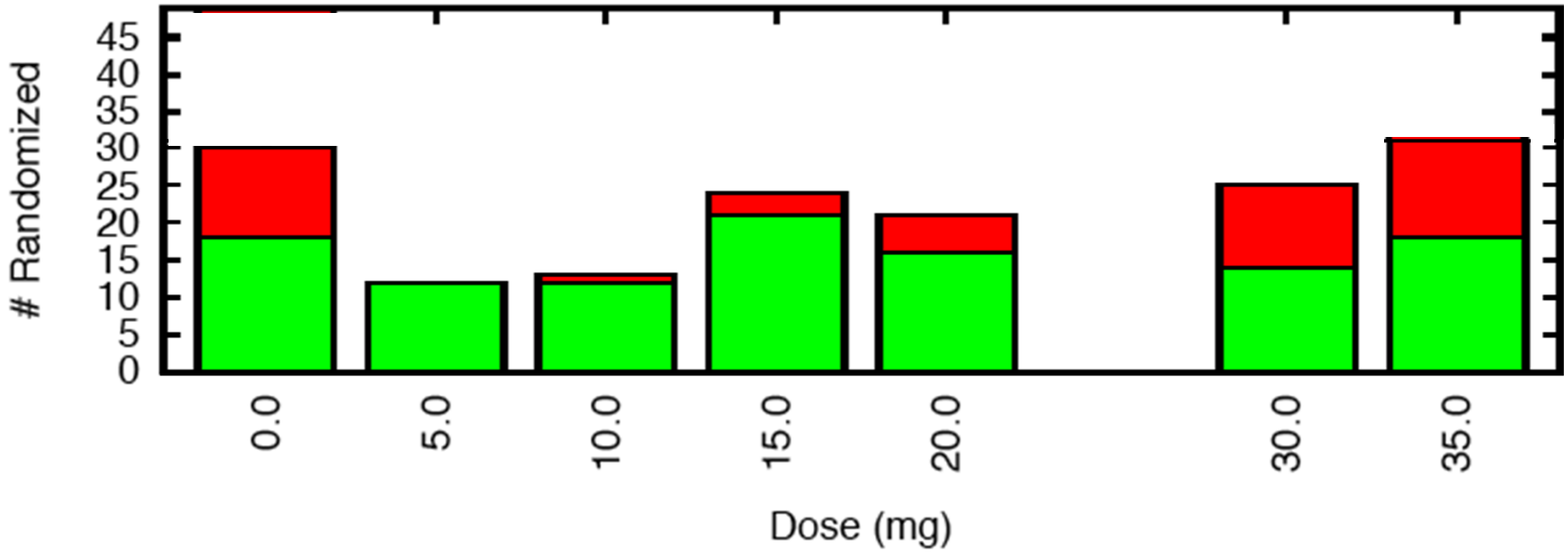

Randomized
45
40
35
30
25
20
15
10
5
0
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

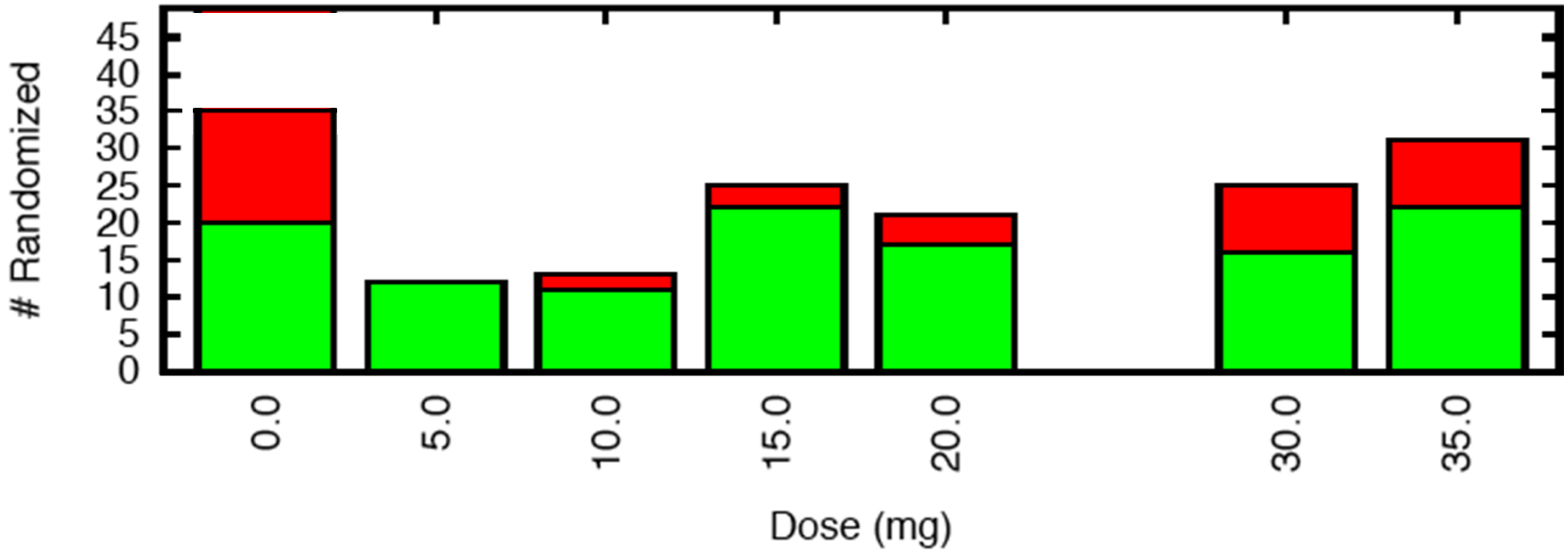

Randomized
45
40
35
30
25
20
15
10
5
0
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

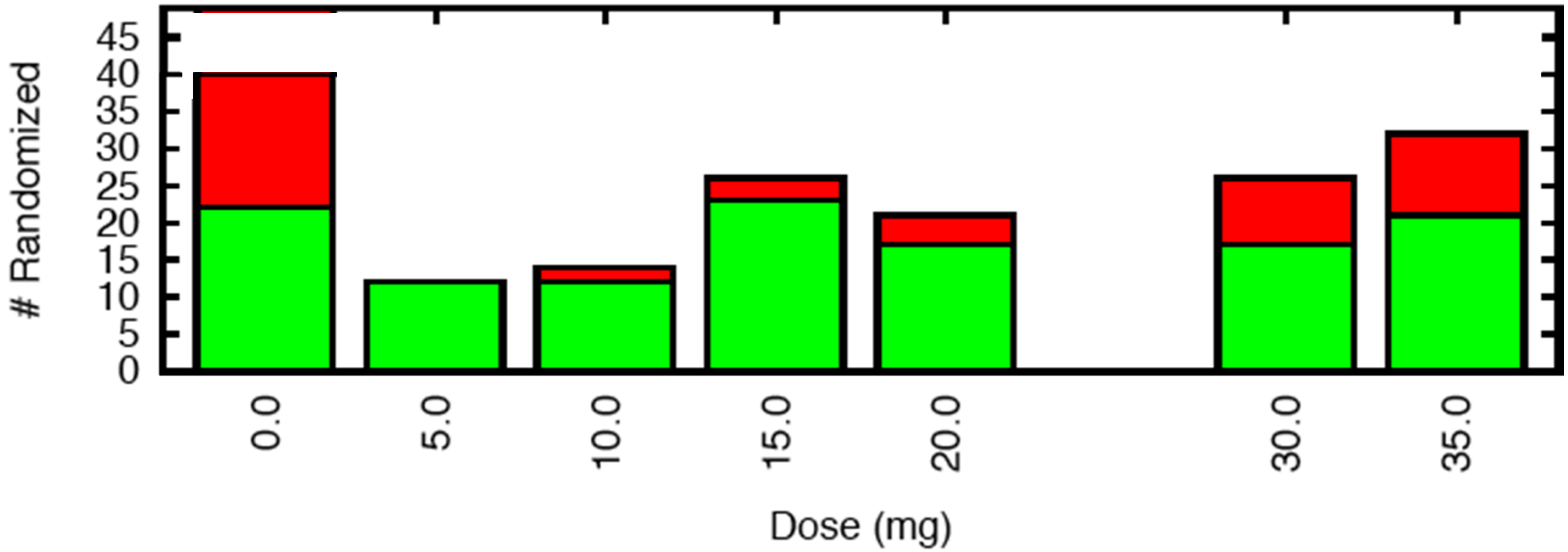

Randomized
45
40
35
30
25
20
15
10
5
0
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

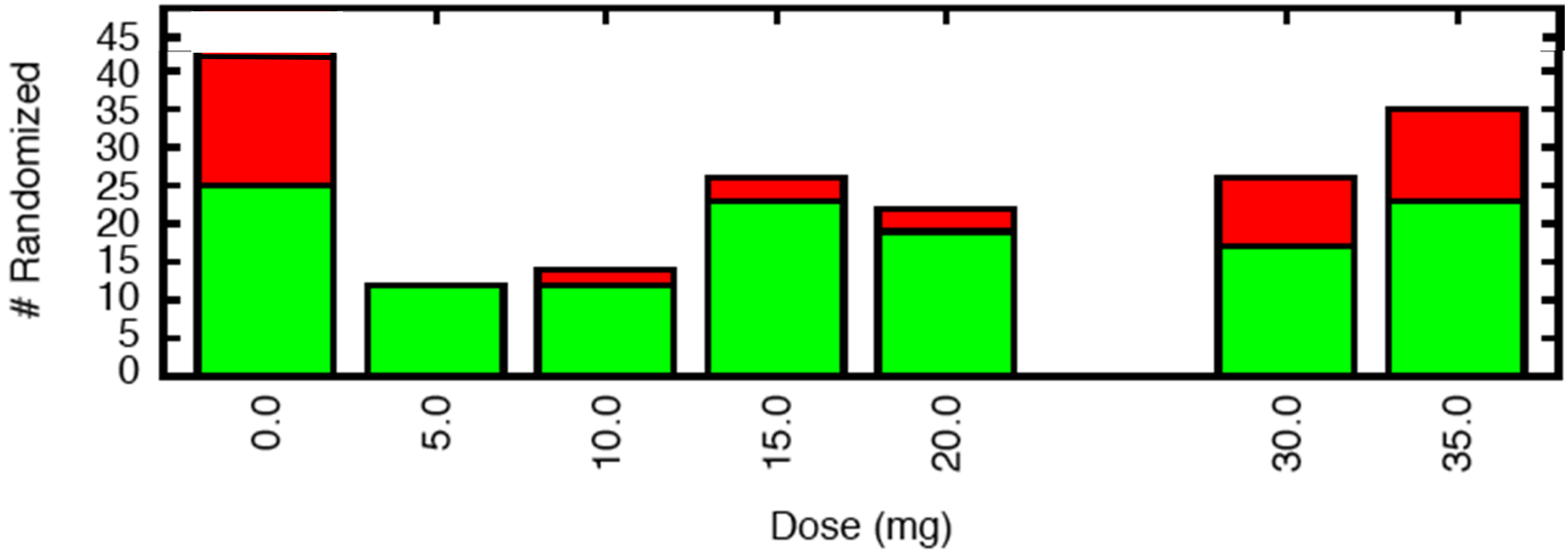

Randomized
45
40
35
30
25
20
15
10
5
0
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

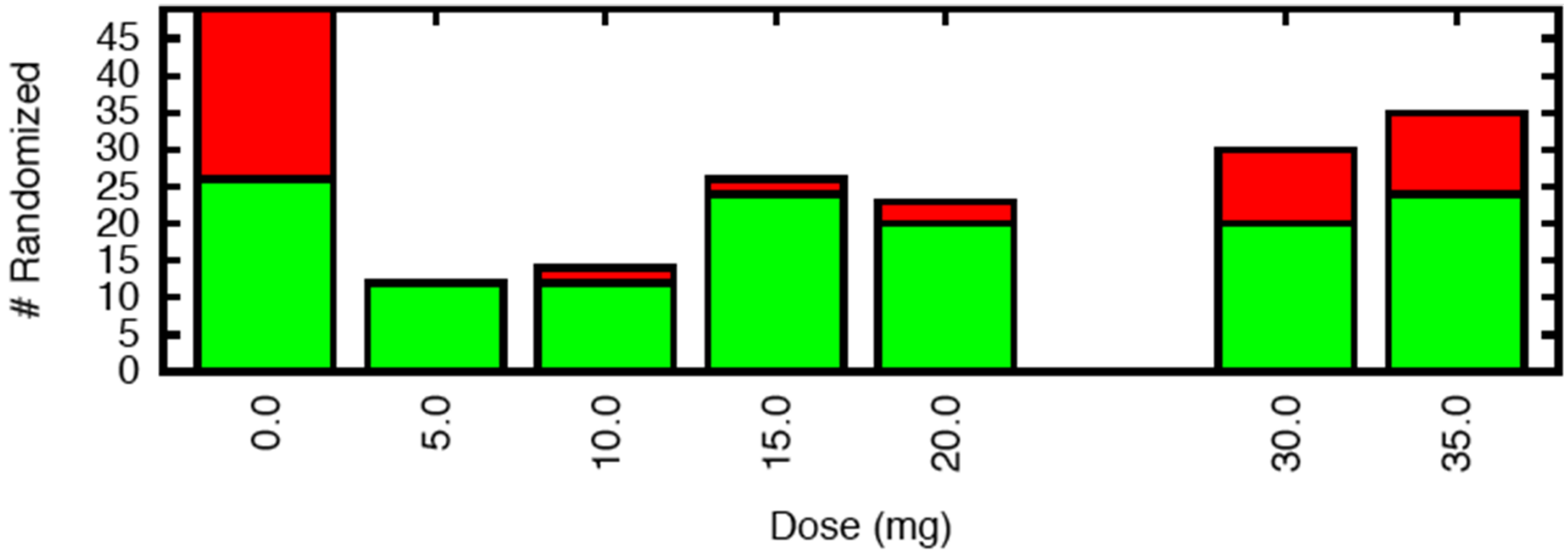

Randomized
45
40
35
30
25
20
15
10
5
0
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

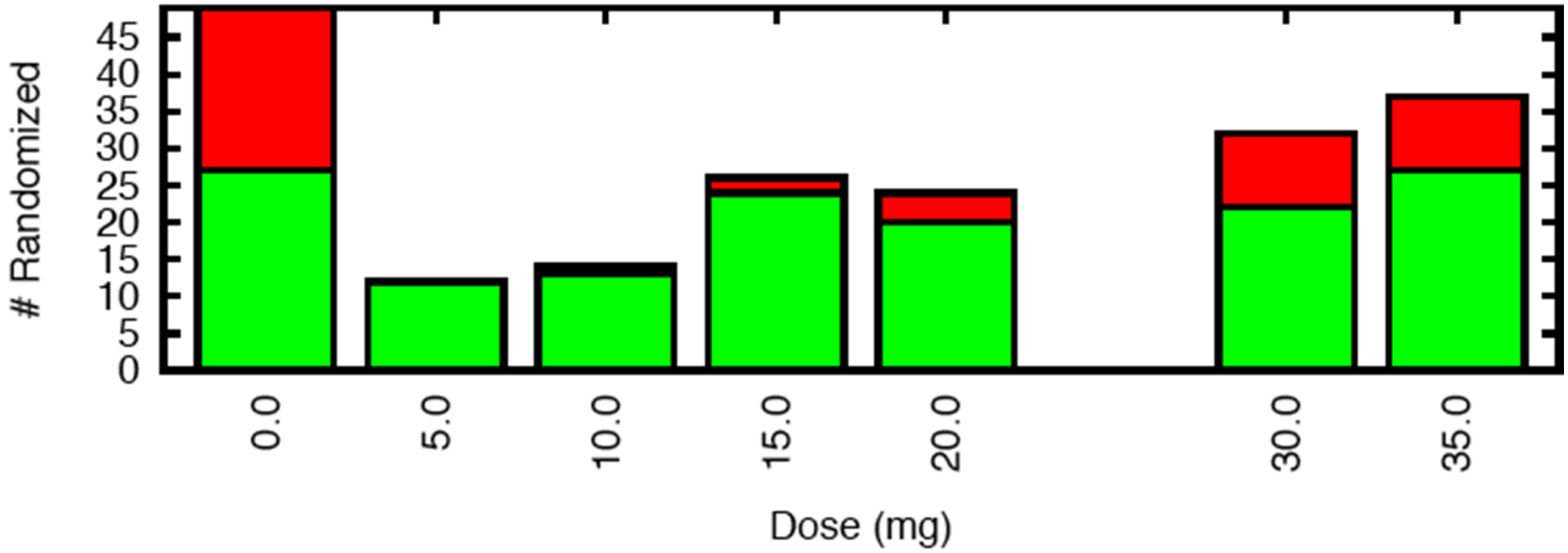

Randomized
45
40
35
30
25
20
15
10
5
0
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

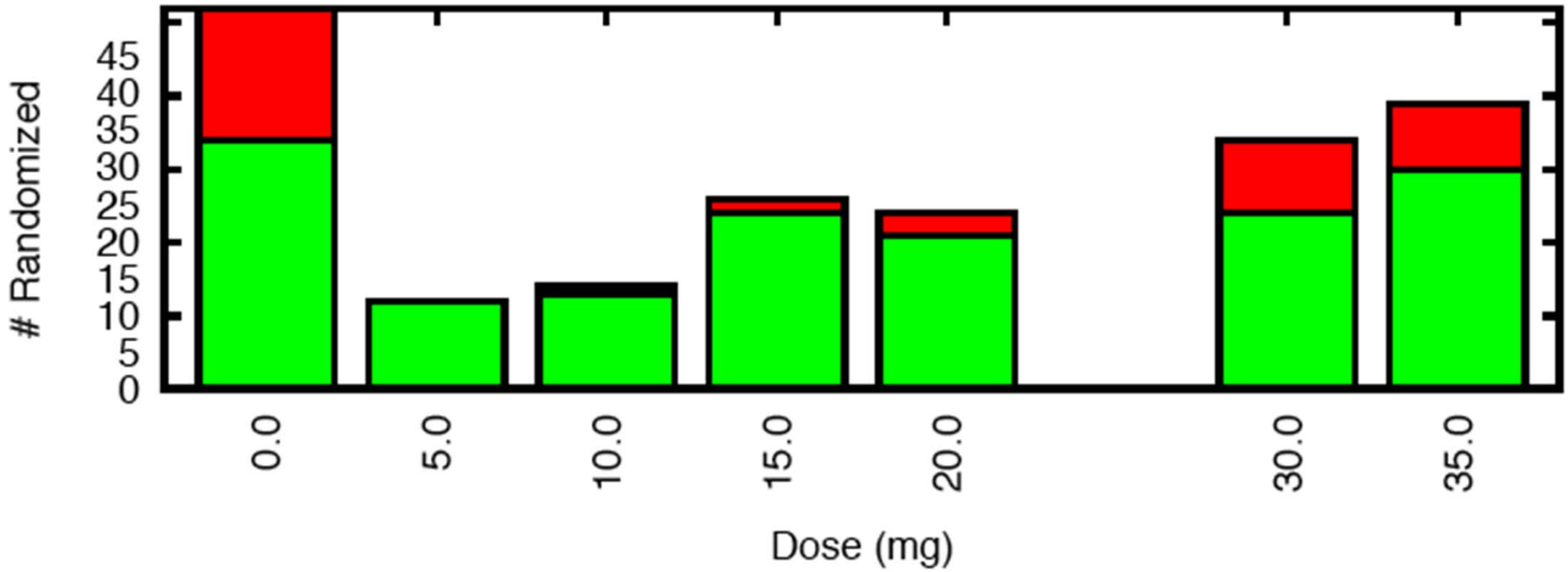

Randomized
45
40
35
30
25
20
15
10
5
0
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

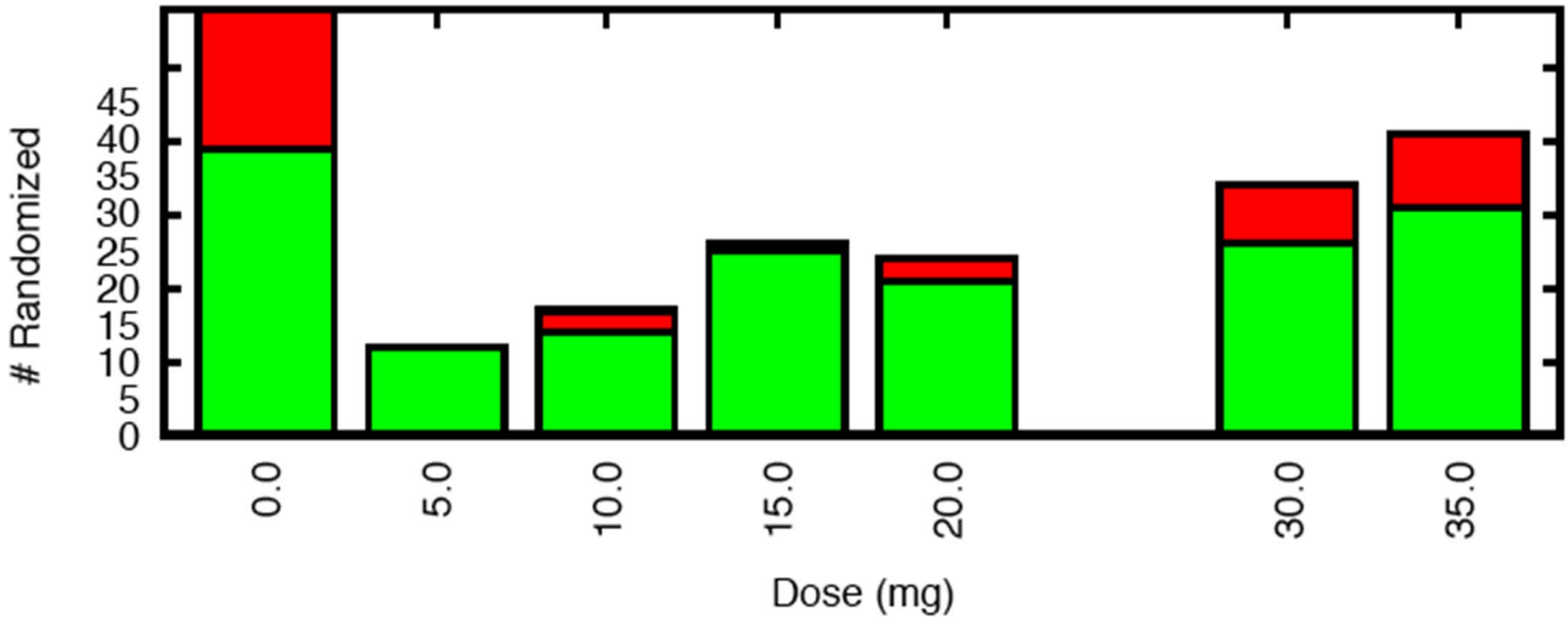

Randomized
45
40
35
30
25
20
15
10
5
0
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

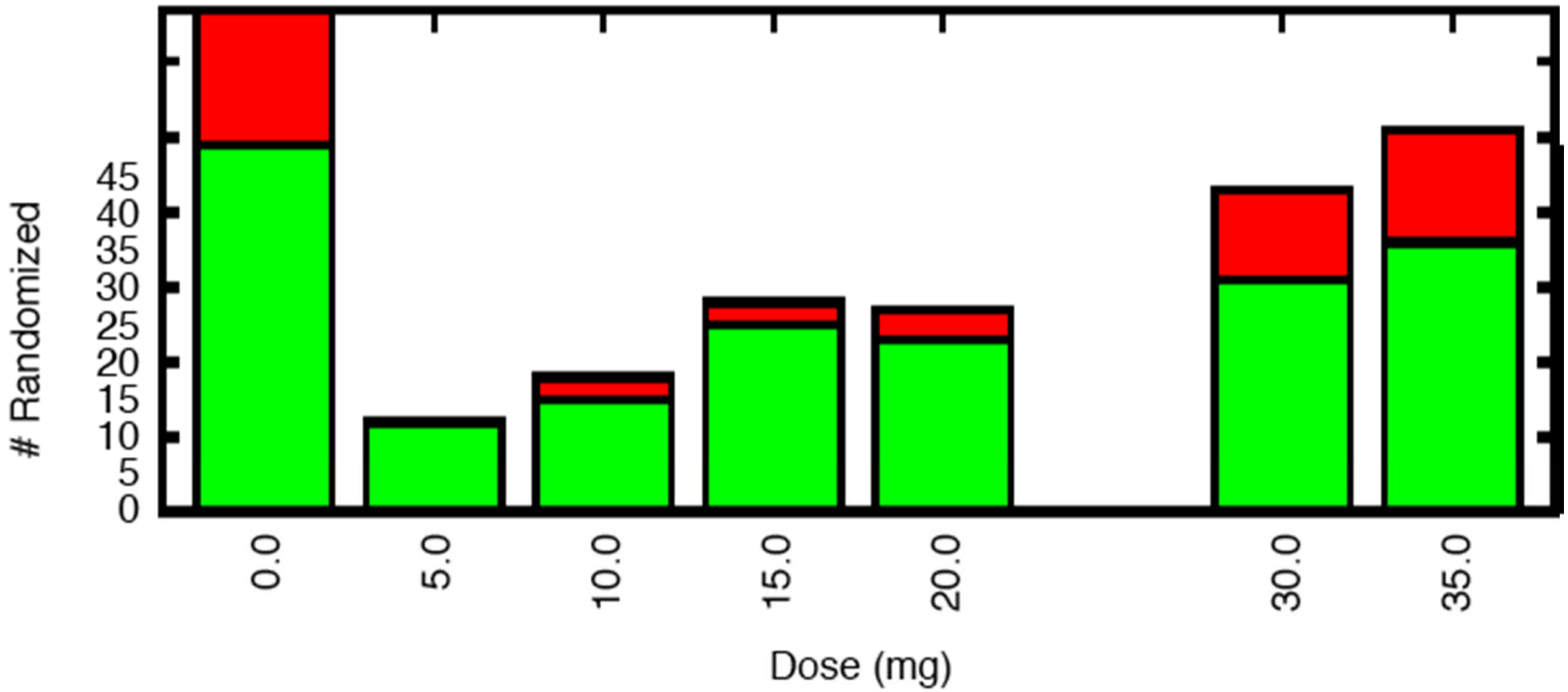

Randomized
45
40
35
30
25
20
15
10
5
0
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

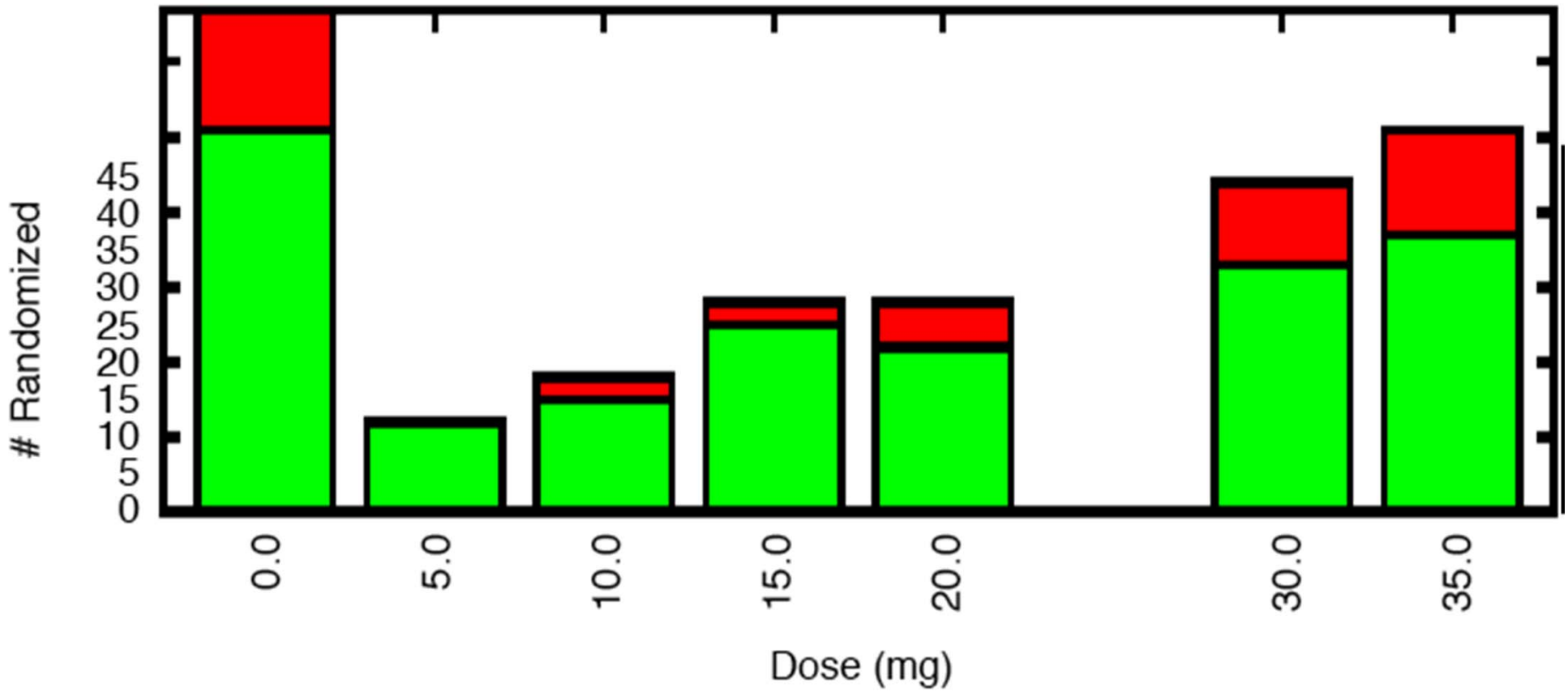

Randomized
45
40
35
30
25
20
15
10
5
0
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

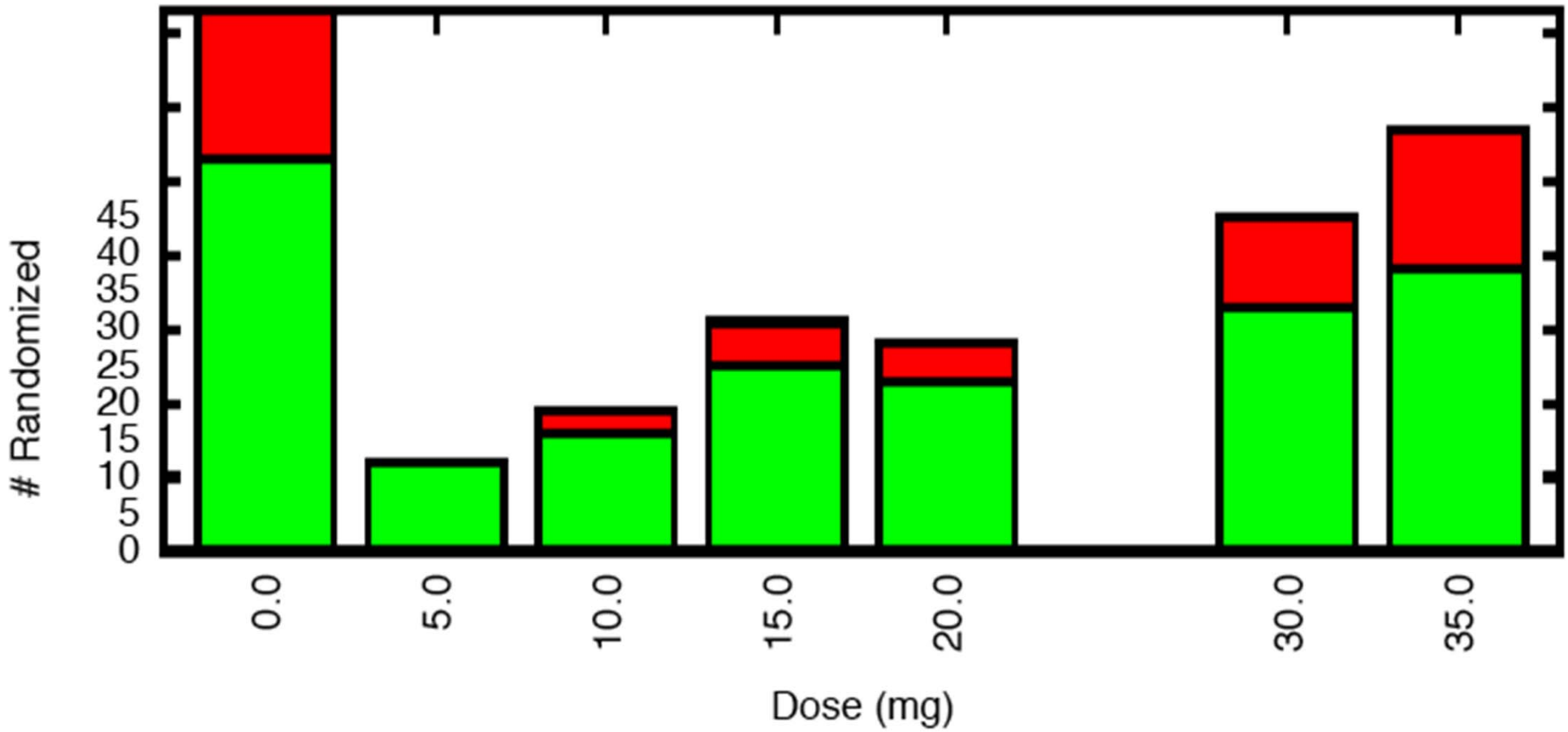

Randomized
45
40
35
30
25
20
15
10
5
0
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

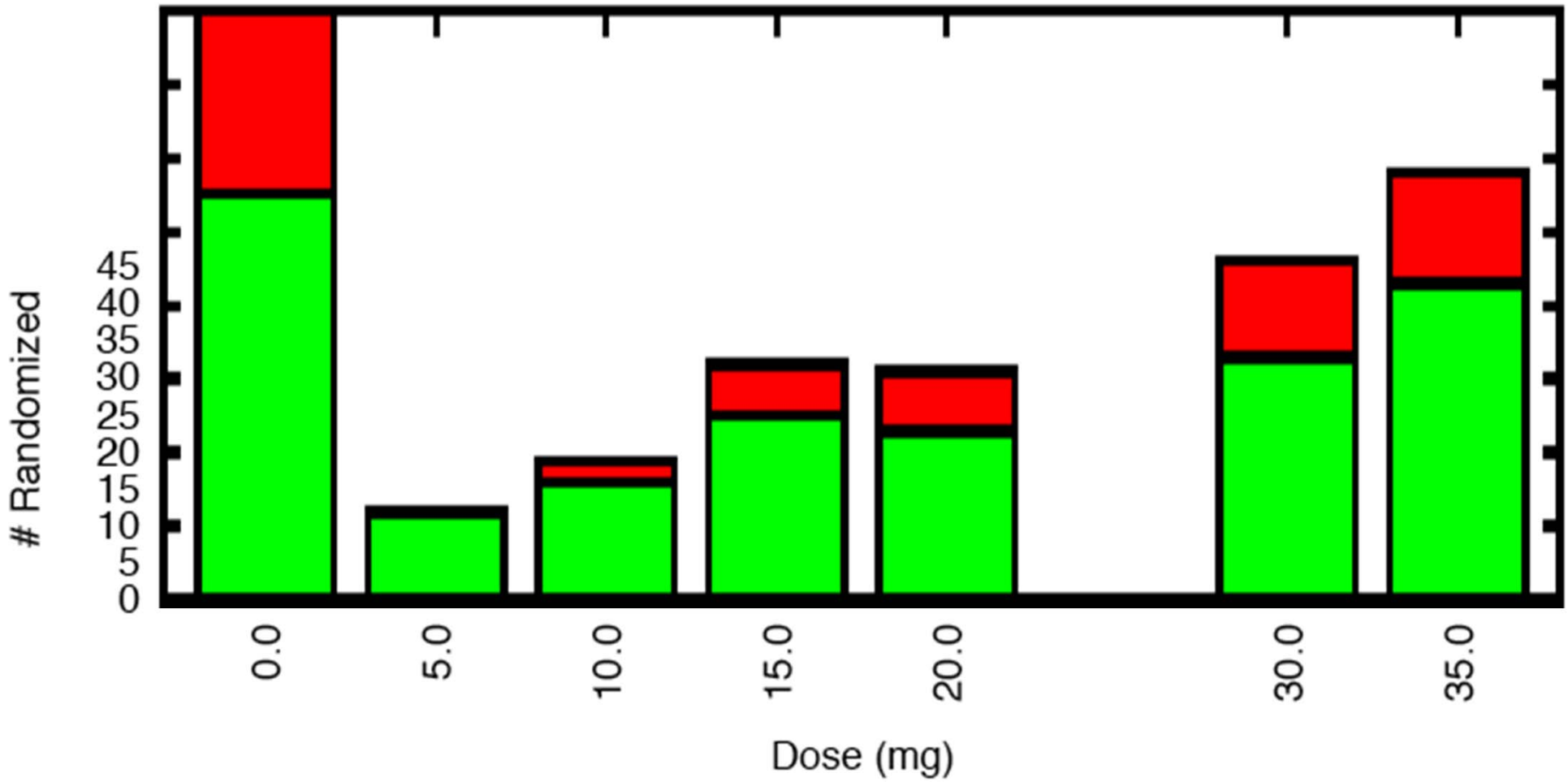

Randomized
45
40
35
30
25
20
15
10
5
0
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

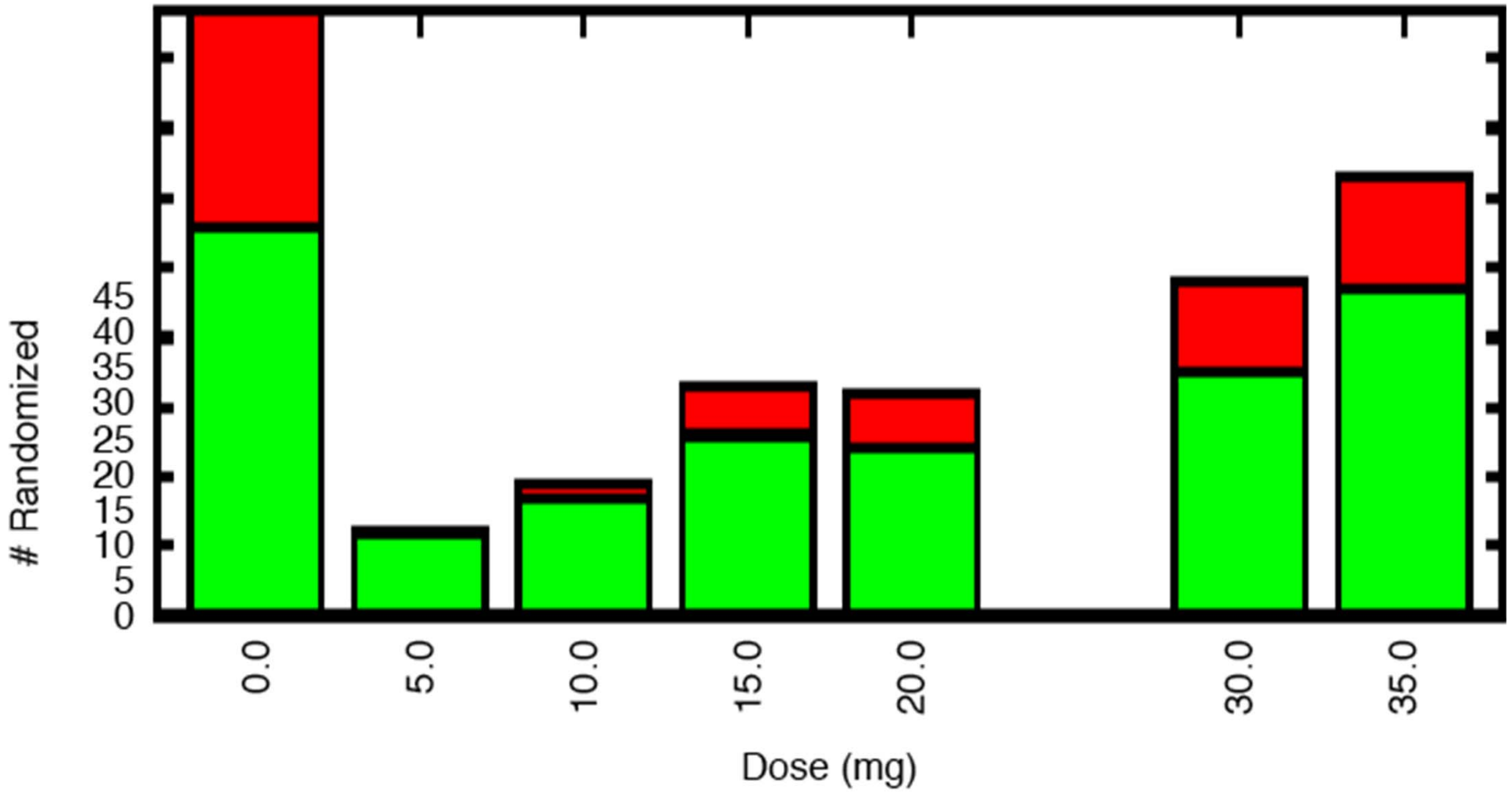

Randomized
0
5
10
15
20
25
30
35
40
45
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

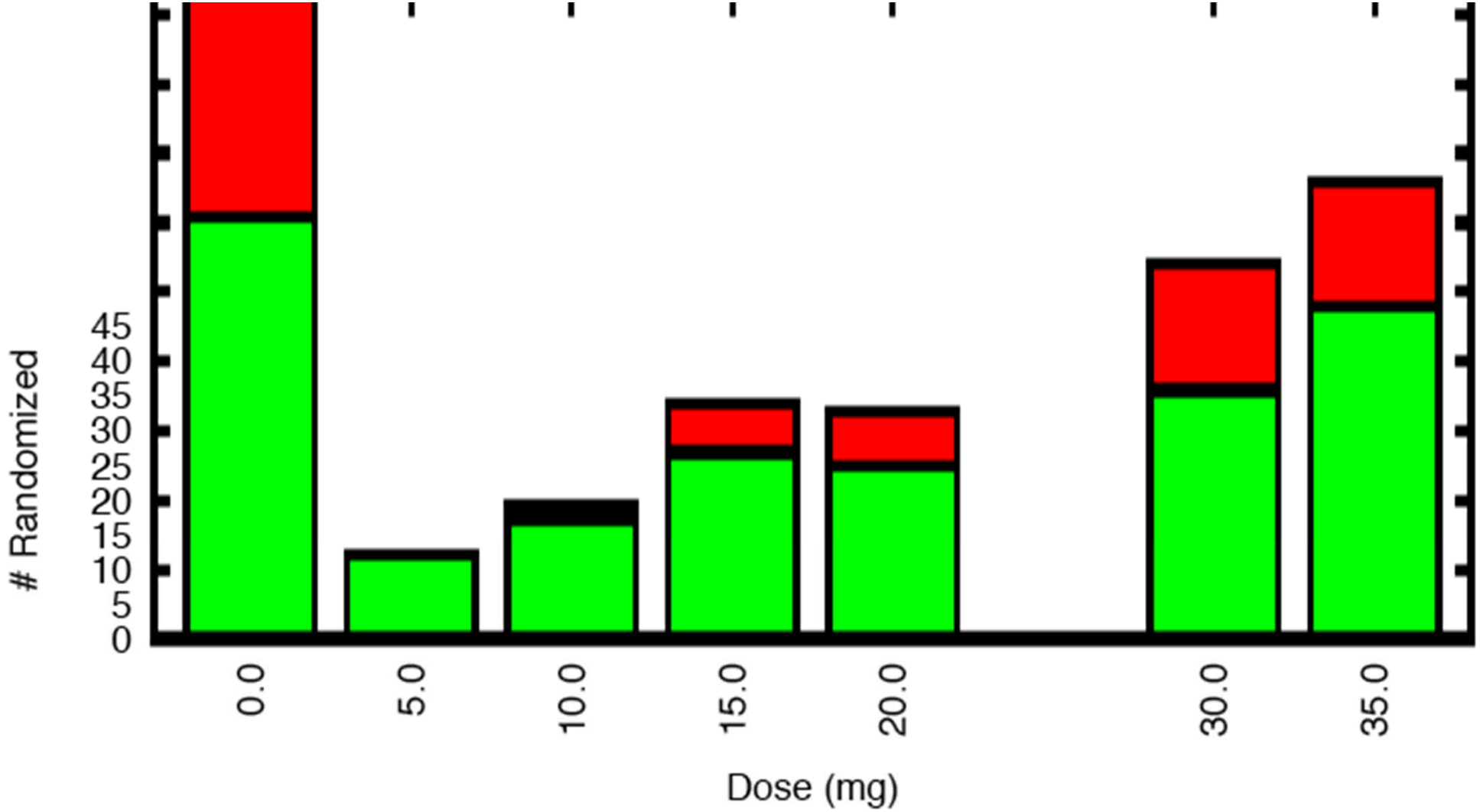

Randomized
0
5
10
15
20
25
30
35
40
45
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

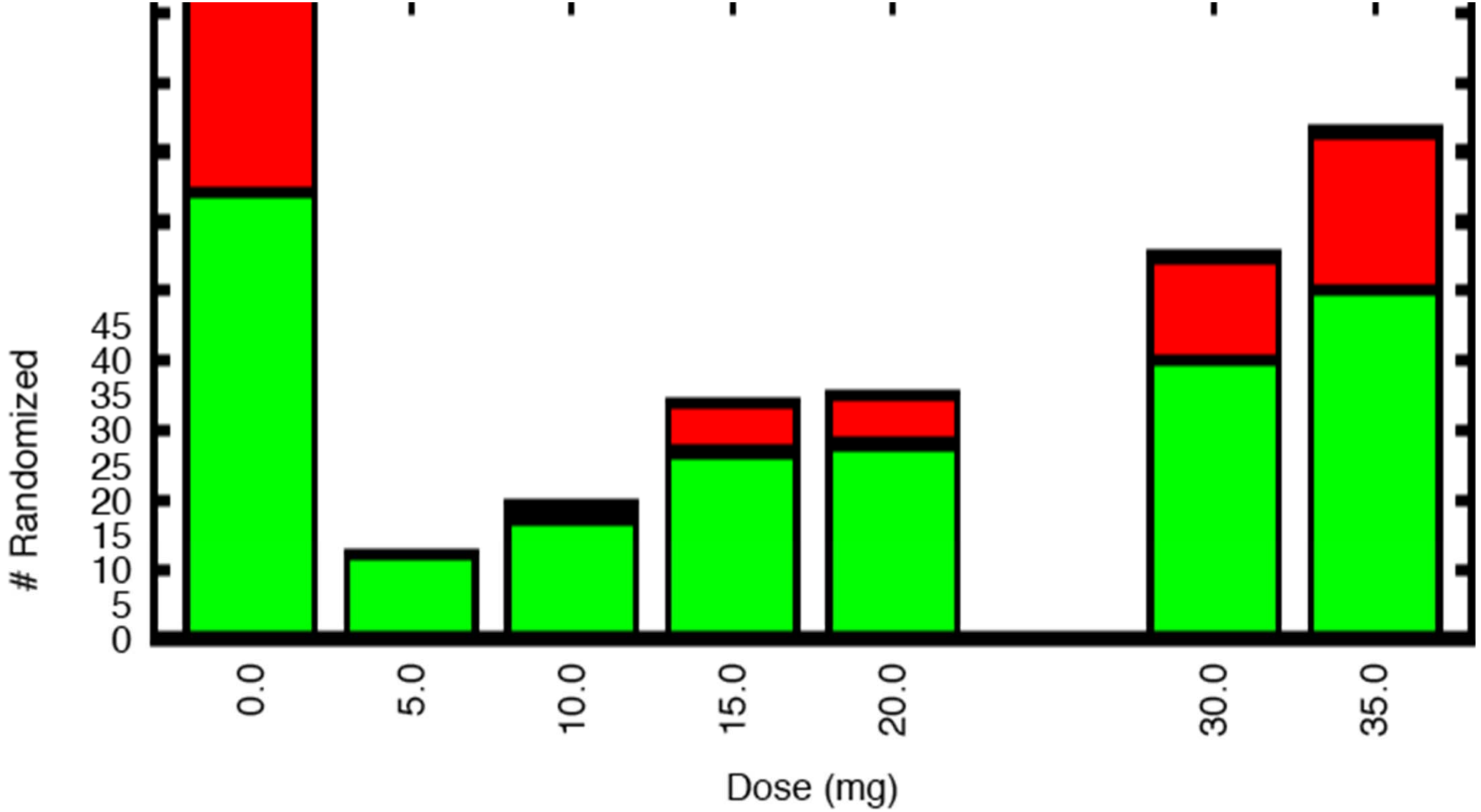

Randomized
0
5
10
15
20
25
30
35
40
45
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)


Randomized
0
5
10
15
20
25
30
35
40
45
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

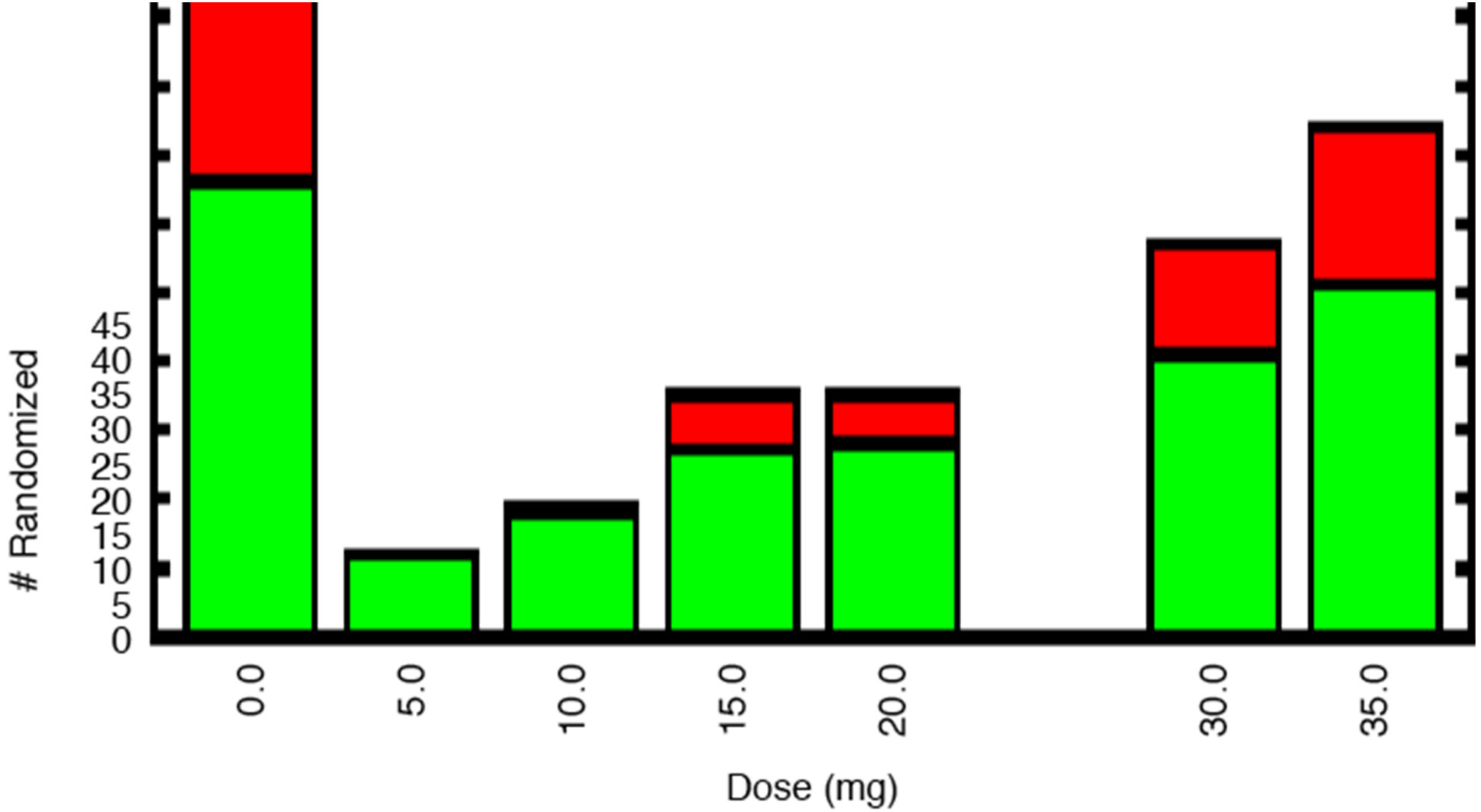

Randomized
0
5
10
15
20
25
30
35
40
45
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)


Randomized
0
5
10
15
20
25
30
35
40
45
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)


Randomized
0
5
10
15
20
25
30
35
40
45
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

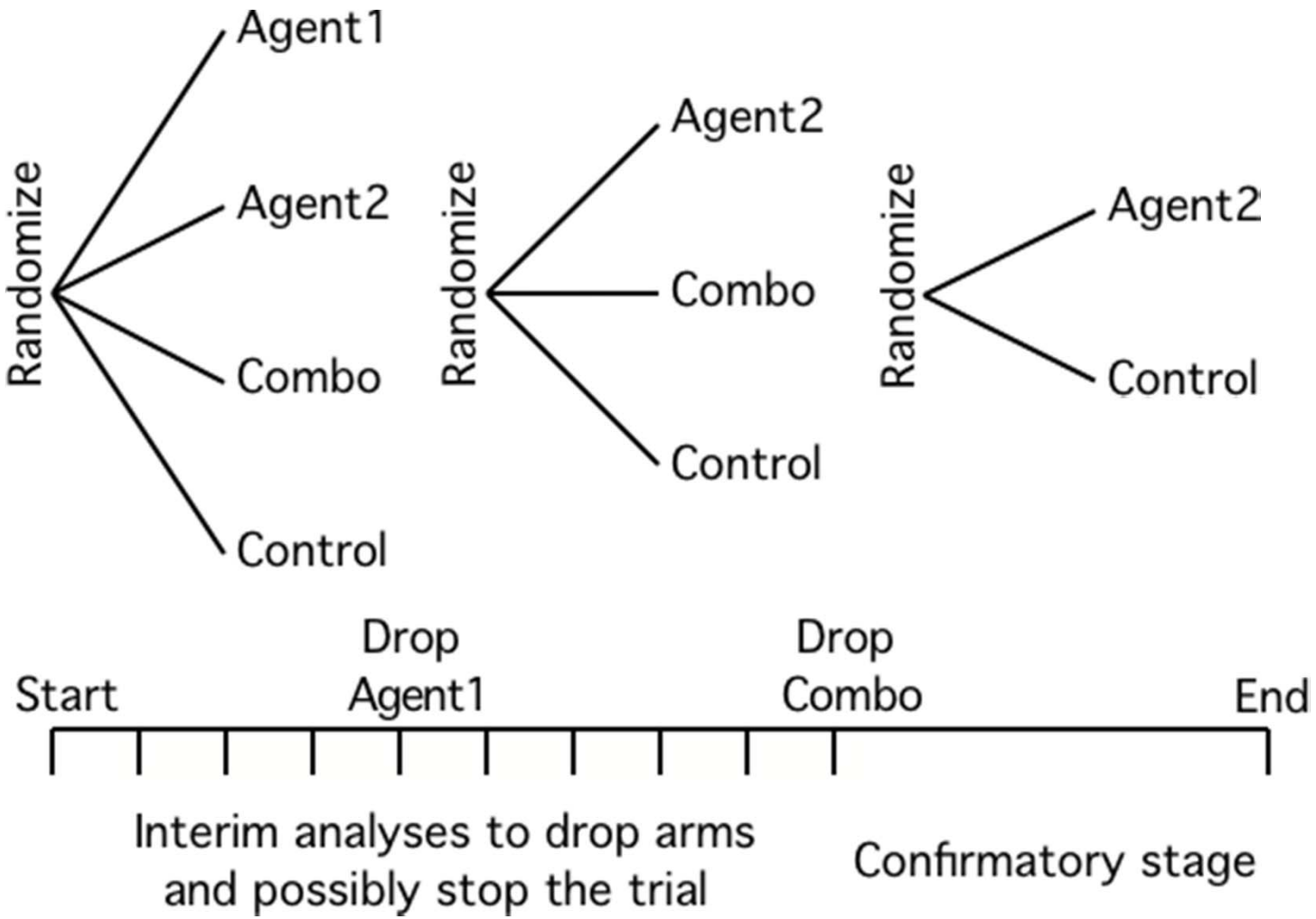


May 14, 2009: Stop for futility


Randomized
0
5
10
15
20
25
30
35
40
45
0.0
5.0
10.0
15.0
20.0
30.0
35.0
Dose (mg)

Final

Adaptive Phase II/III Trial



IoM Report on Cancer Cooperative Groups 2010

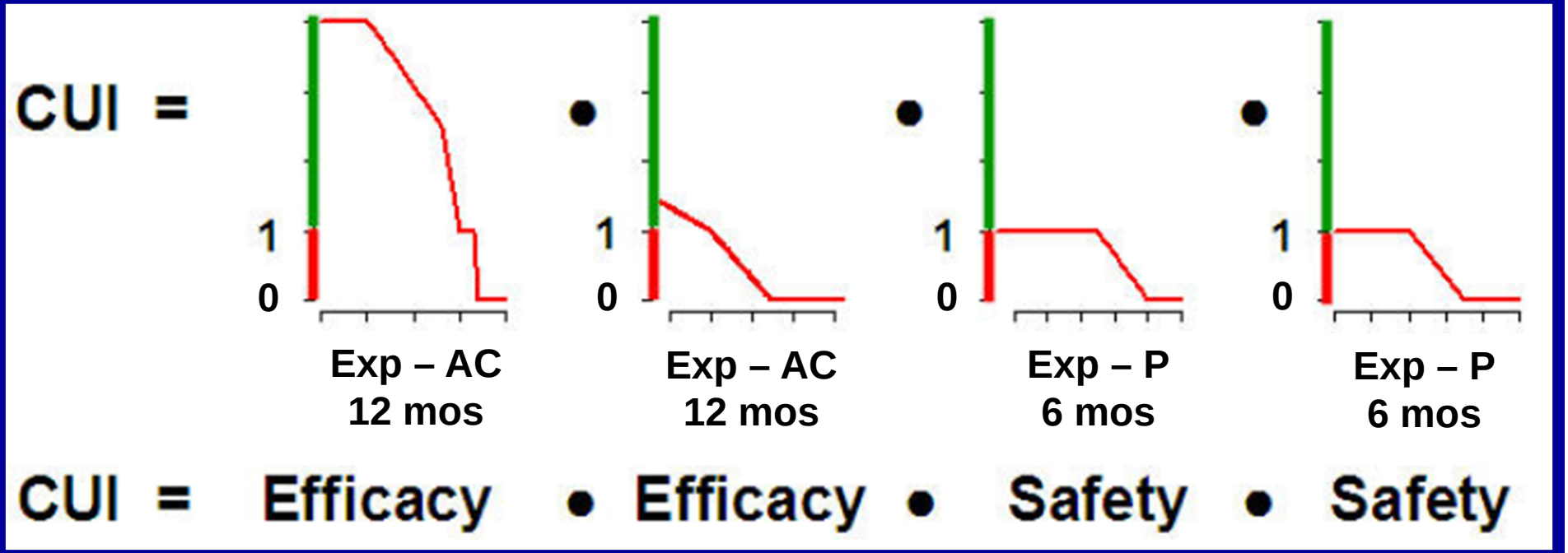
Example from Critical Path Initiative

- Type II diabetes
- Seamless Phase II/III: Dose finding plus confirmation
- Sample size 200 - 1566
- Active comparator & placebo
- Primary endpoint: Clinical Utility Index (12 months)

Some Details

- Phase II: 7 doses experimental drug plus placebo plus active control
- Phase III
  - 1 or 2 doses experimental drug
  - Sample size via predictive power considering available Phase II data
  - Adaptive transition: Bayesian predictive probs
- Both phases driven by CUI

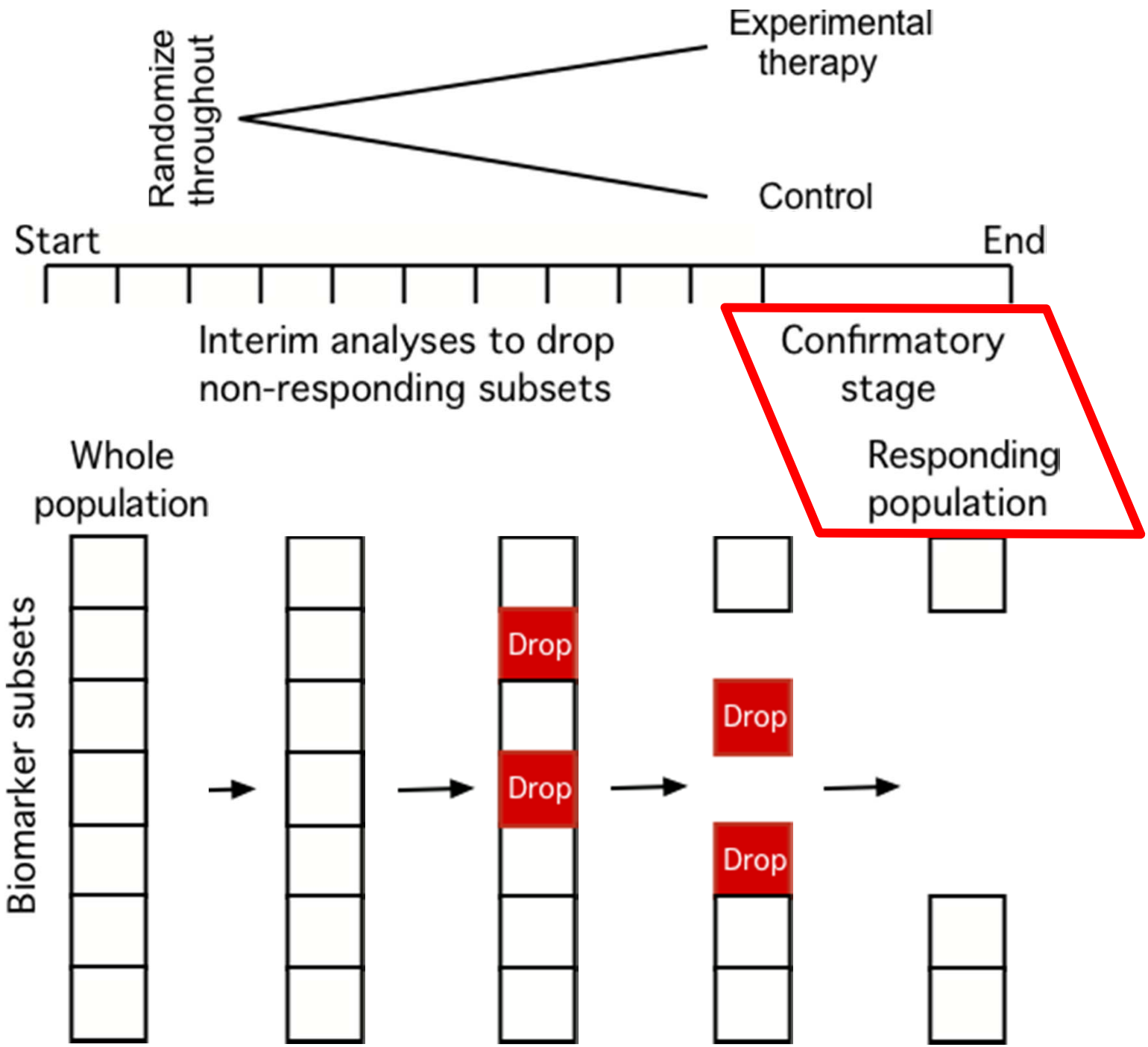
Clinical Utility Index



- Dose-response modeling
- Longitudinal modeling

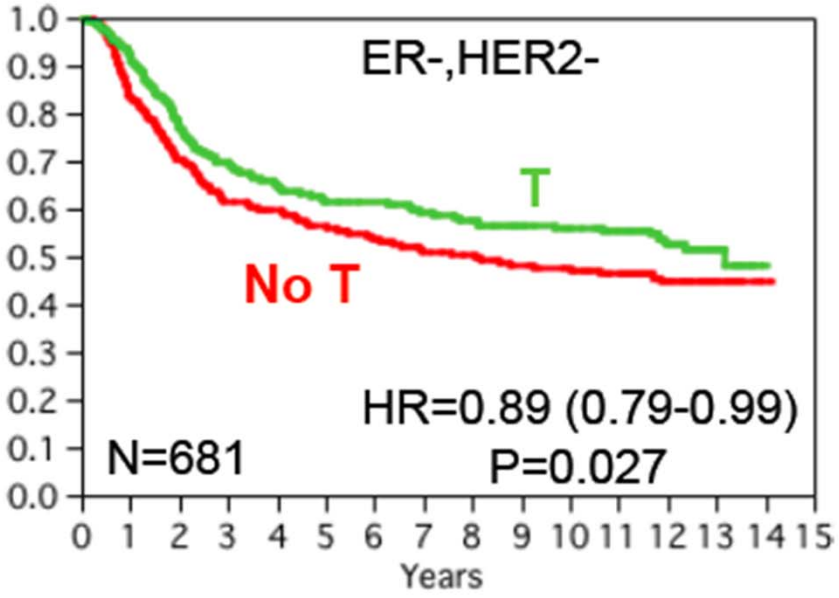
Adaptive Biomarker Trial

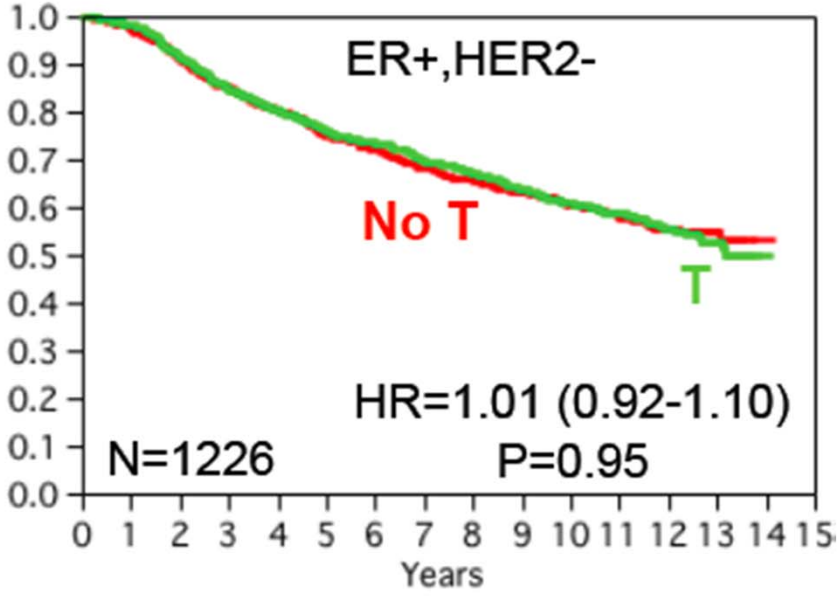
EXAMPLE TRIAL

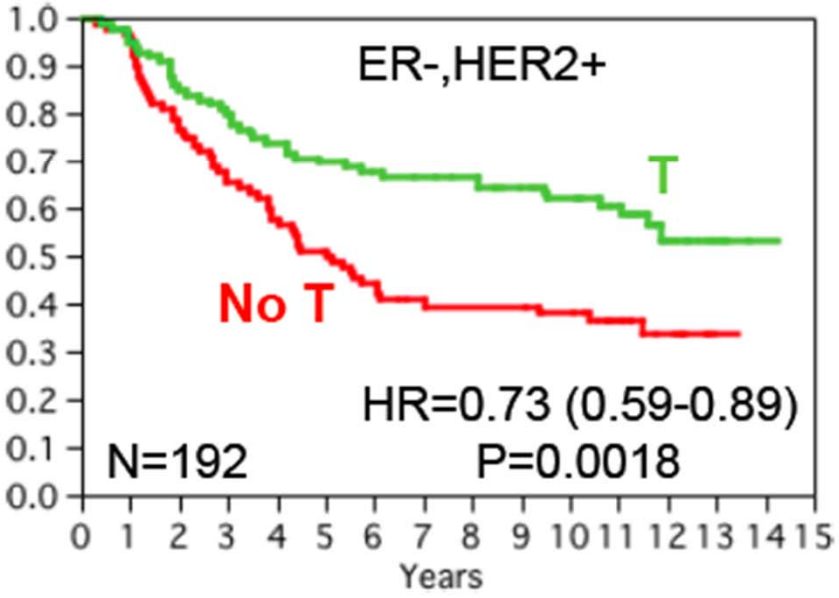


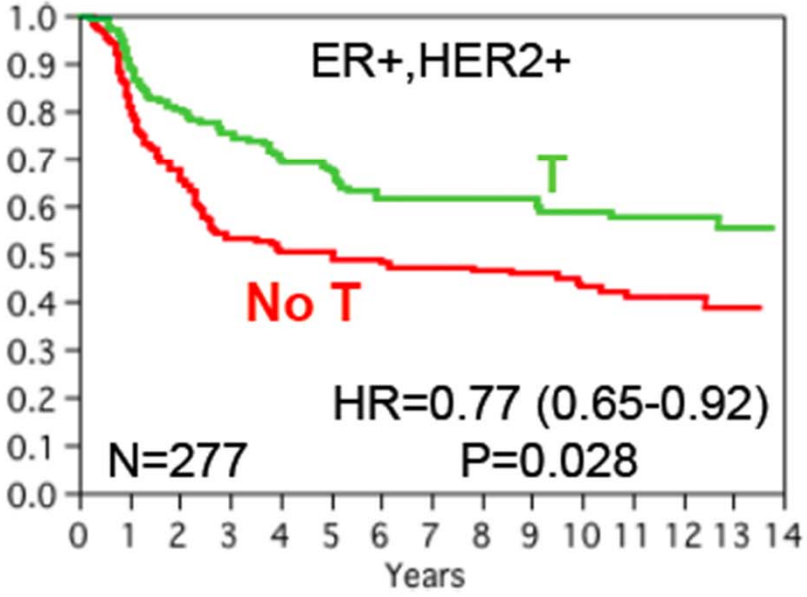
Savings possible in sample size when using biomarkers …

Relapse-free survival in CALGB 9344; n = 2376









Berry et al. SABCS 2009

THE WALL STREET JOURNAL.
WSJ.com

THE SATURDAY ESSAY | OCT

A New Rx for Med

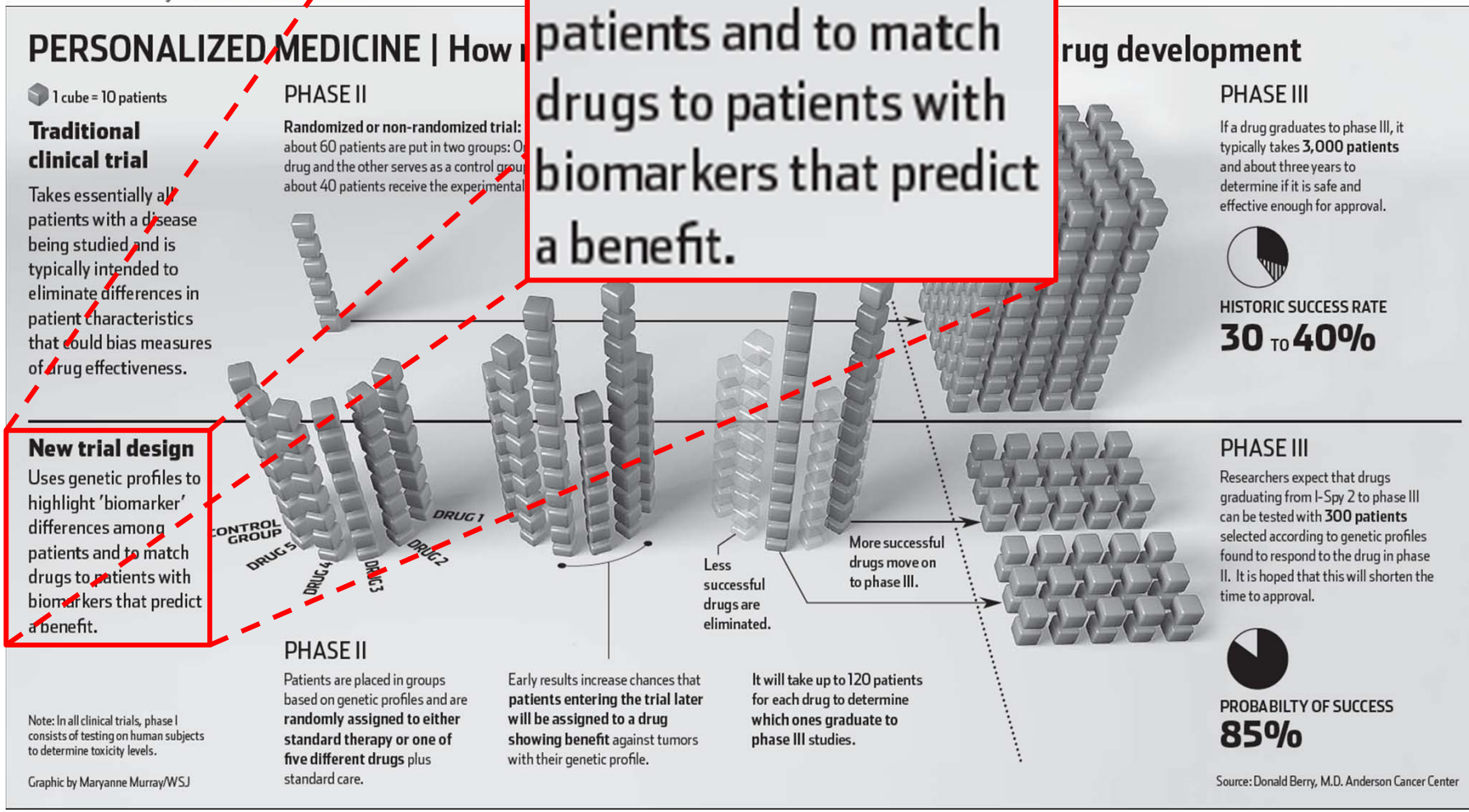
Fed up with slow drug trials, ca ack to personalized treatments.

By RON WINSLOW

New trial design

Uses genetic profiles to highlight 'biomarker' differences among patients and to match drugs to patients with biomarkers that predict a benefit.

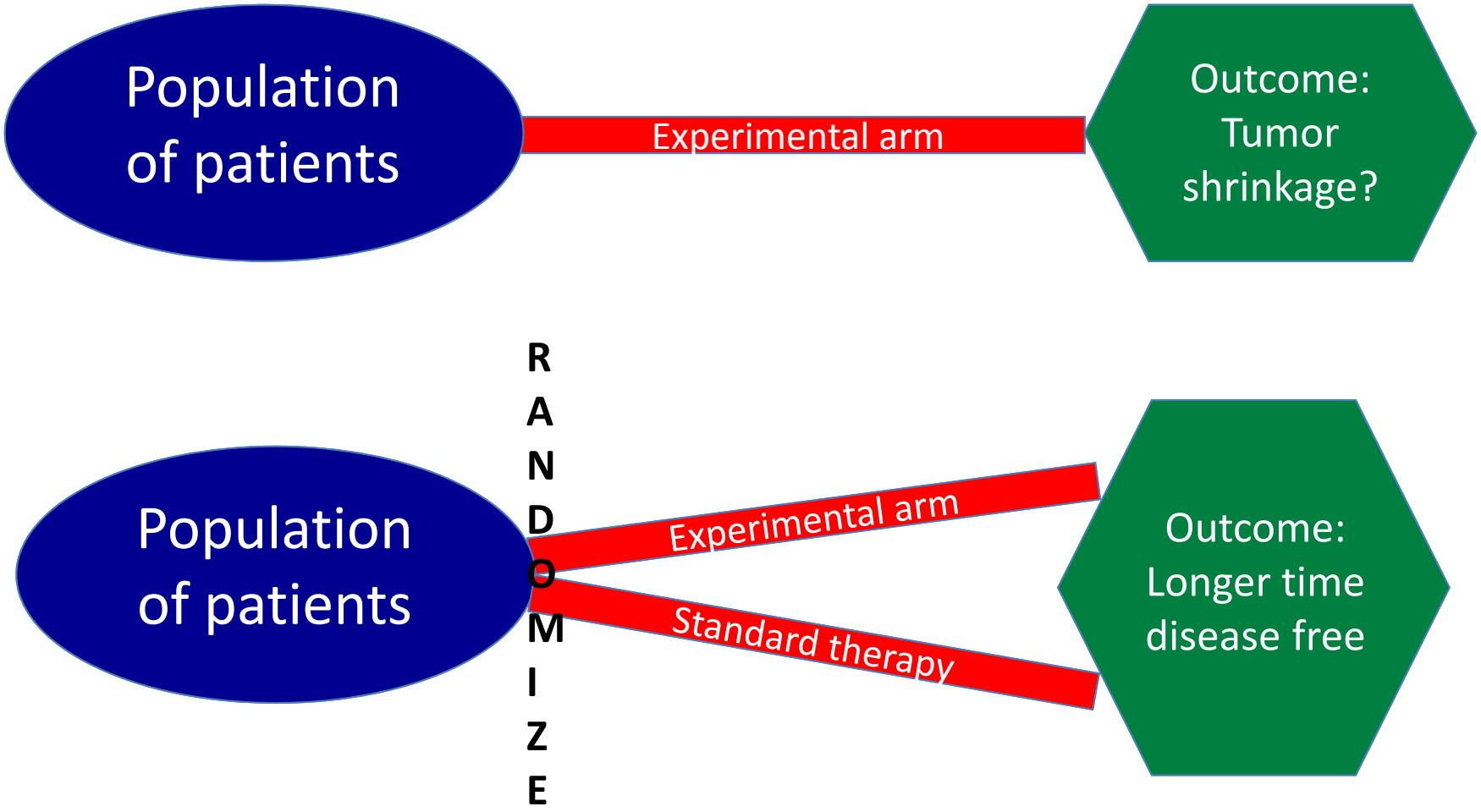
PERSONALIZED MEDICINE | How rug development



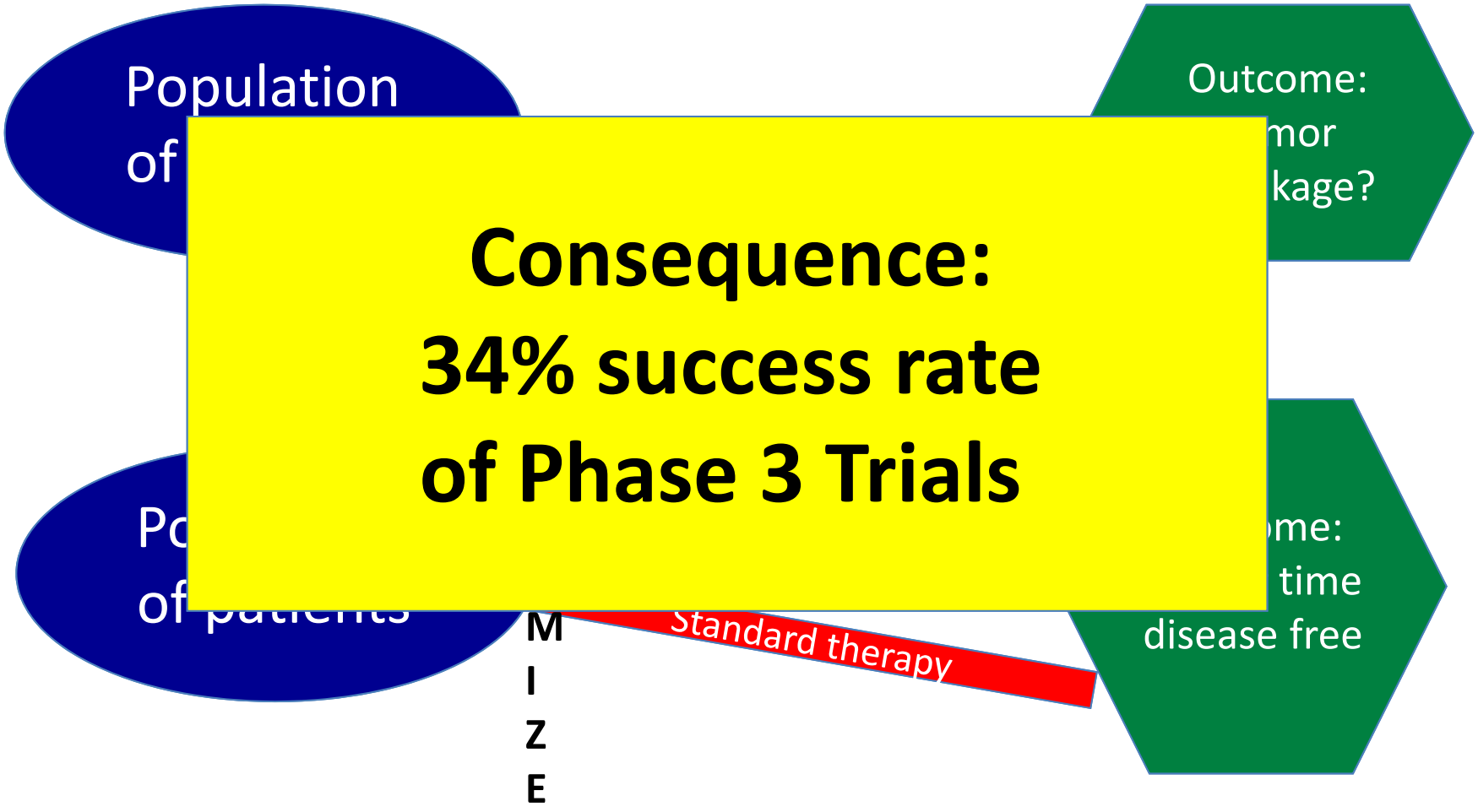
I-SPY2: The Cartoon
(Press conference* slides)

*<http://ispy2.org>

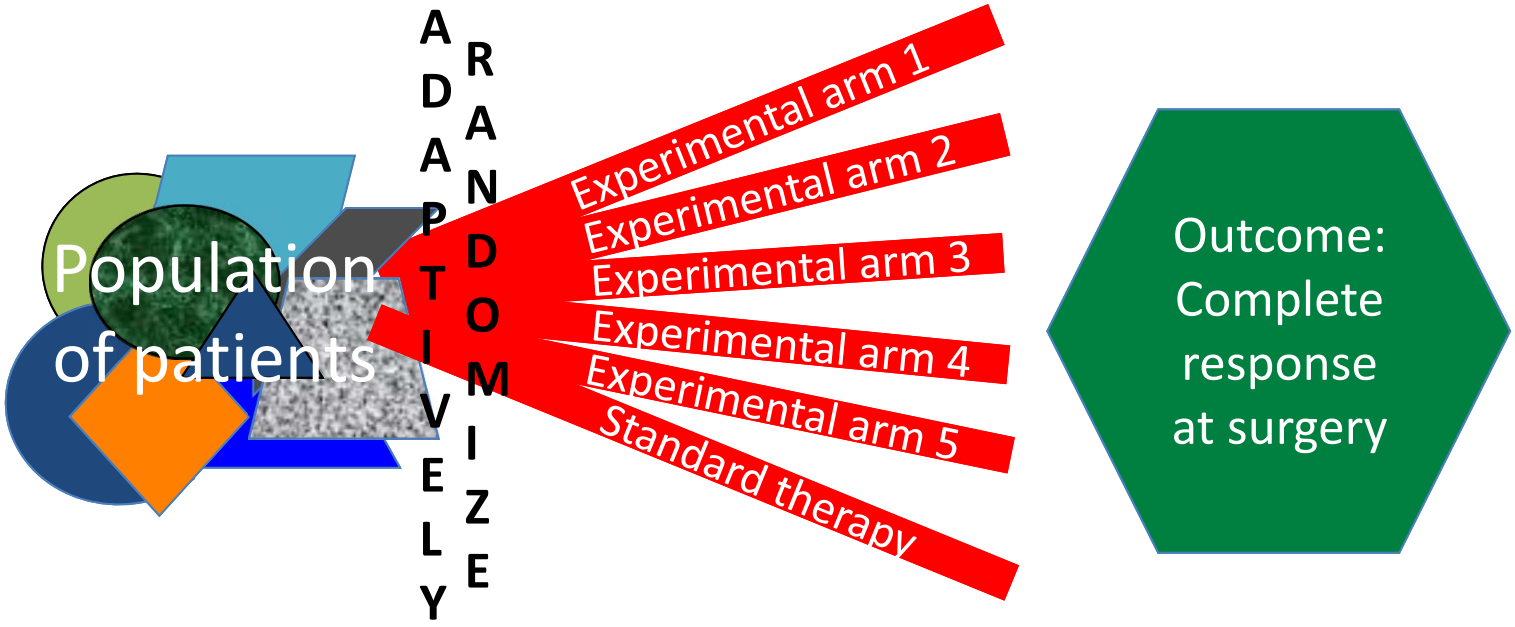
Standard Phase 2 Cancer Drug Trials



Standard Phase 2 Cancer Drug Trials



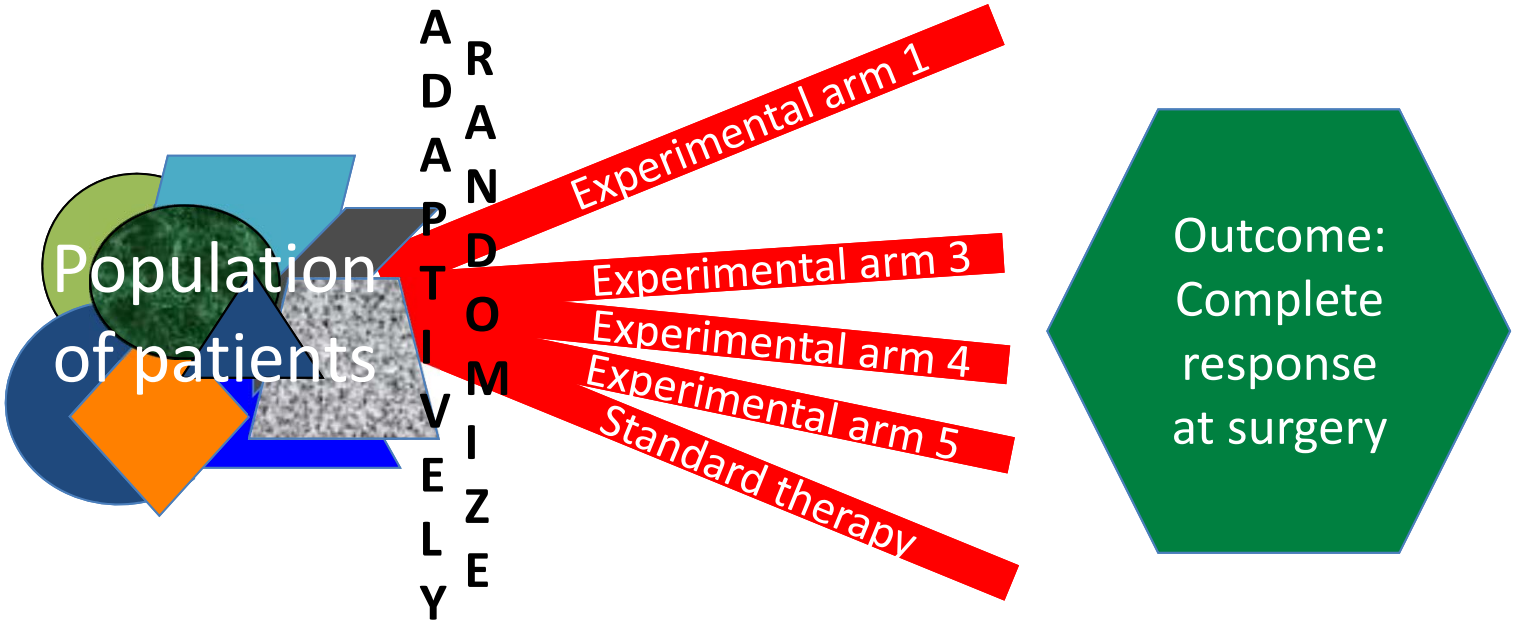
I-SPY2 TRIAL


Population of patients
ADAPTIVELY RANDOMIZE
Experimental arm 1
Experimental arm 2
Experimental arm 3
Experimental arm 4
Experimental arm 5
Standard therapy
Outcome: Complete response at surgery

I-SPY2 TRIAL

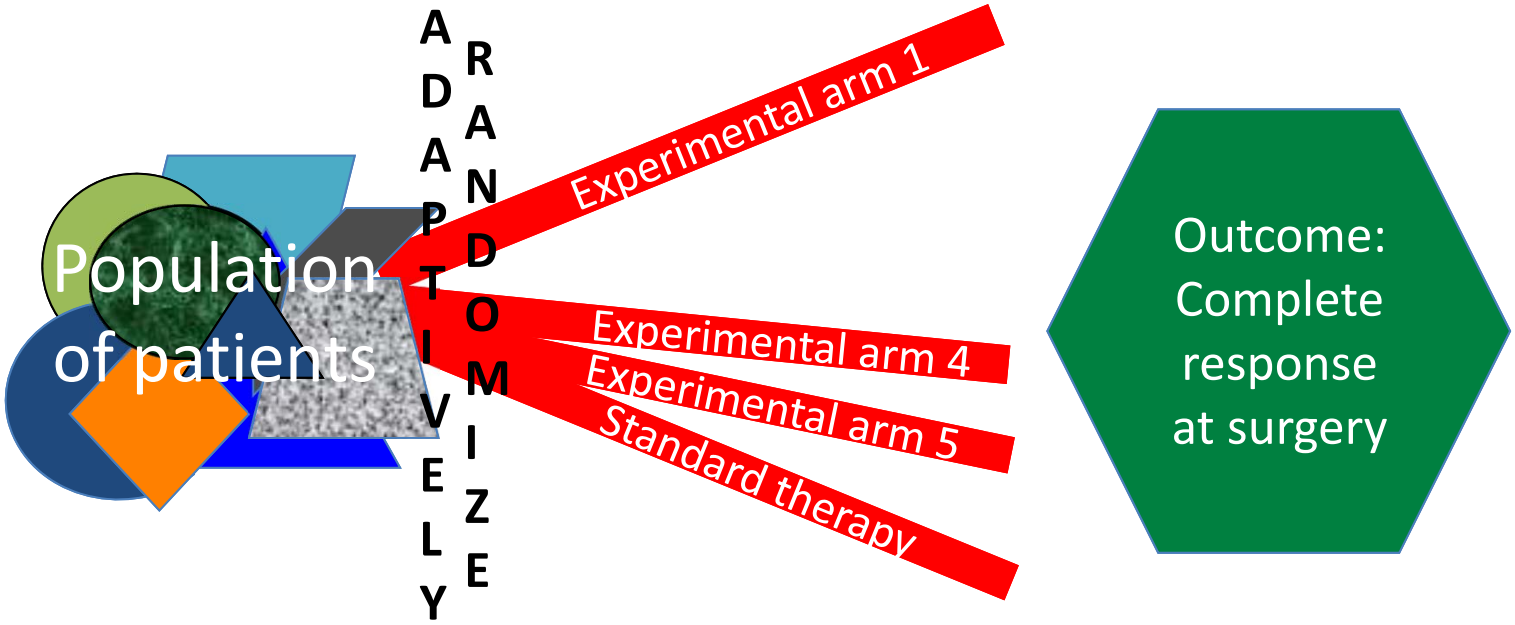
Arm 2 graduates to small focused Phase 3 trial

I-SPY2 TRIAL



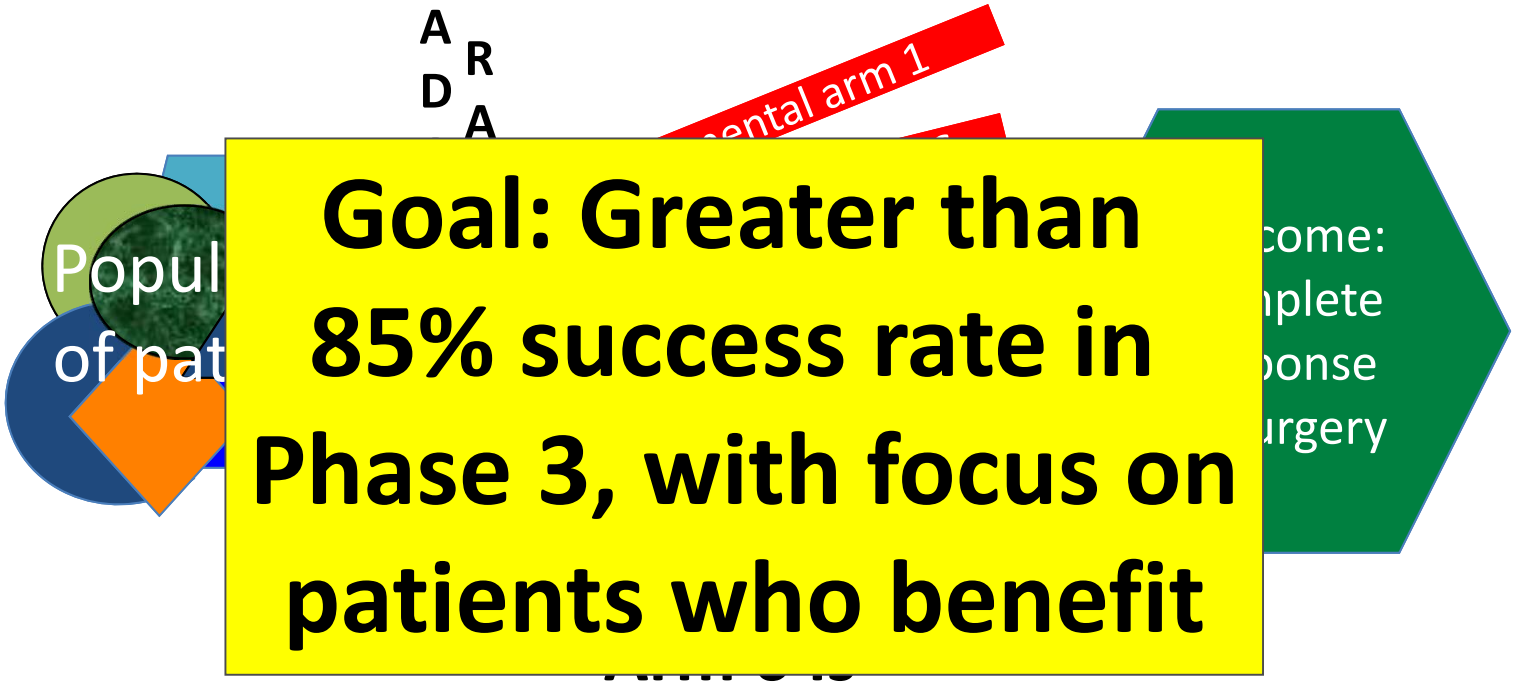
Arm 3 drops for futility

I-SPY2 TRIAL



Arm 5 graduates to small focused Phase 3 trial

I-SPY2 TRIAL



Goal: Greater than 85% success rate in Phase 3, with focus on patients who benefit

added to the mix

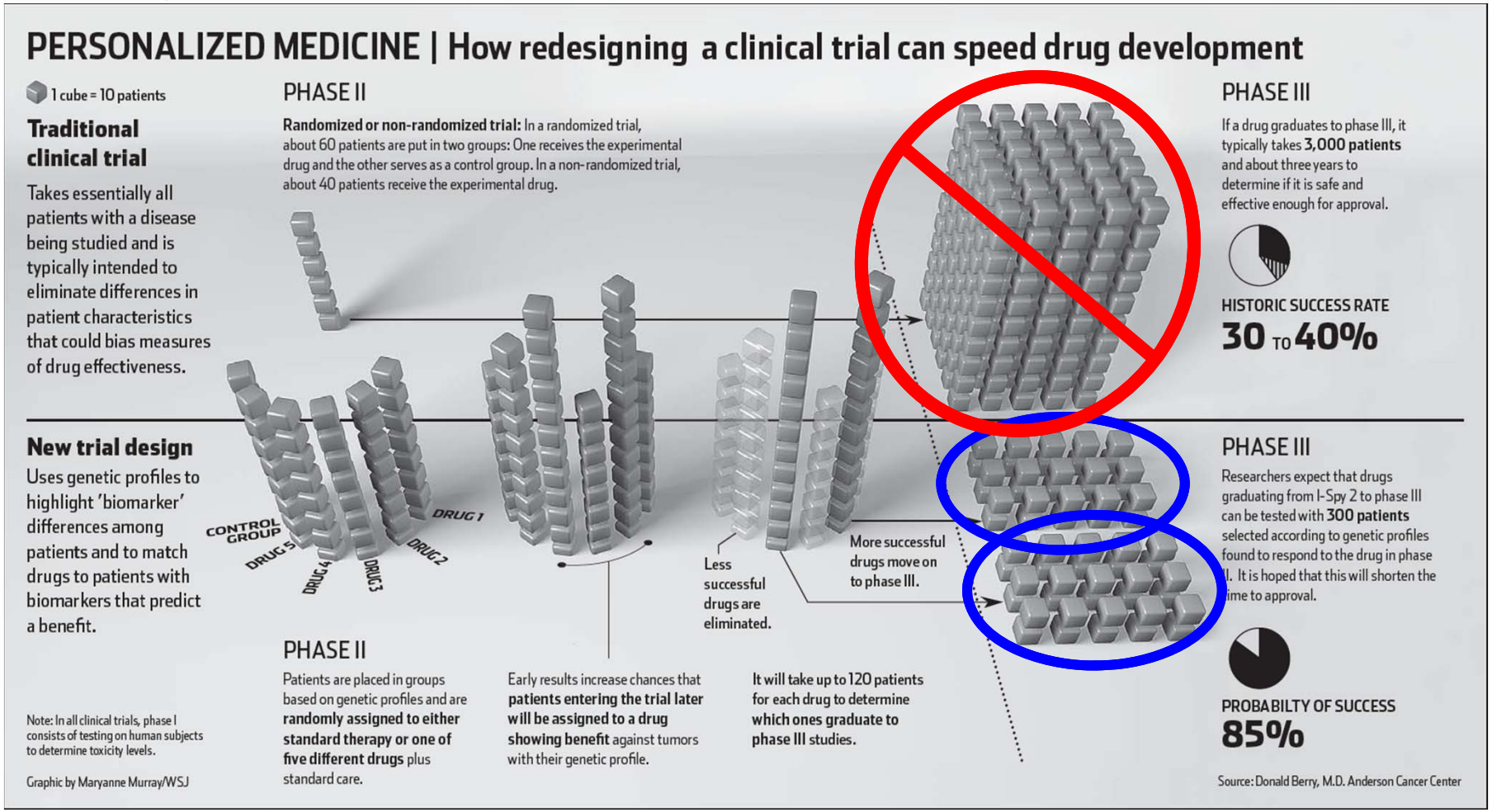
THE WALL STREET JOURNAL.
WSJ.com

THE SATURDAY ESSAY | OCTOBER 2, 2010

A New Rx for Medicine

Fed up with slow drug trials, cancer patients and doctors are testing a fast track to personalized treatments.

By RON WINSLOW



I-SPY2 Adaptive Design Process

- PI: Laura Esserman, UCSF
- Sponsored by FNIH: NCI, FDA, industry, academia, 

- Coordinated with FDA (CDER, CBER & CDRH) from inception
- Current status: 20 centers, ~50 patients, experimental drugs so far: neratinib, ABT888, AMG386

I-SPY 2 Effects

- Match drugs (& combos) with biomarker signatures
- Graduate drug/biomarker pairs to smaller (n < 300), more focused, more successful Phase 3
- Descendents of I-SPY 2 in melanoma, colorectal cancer, Alzheimer's, acute heart failure, …

Software?

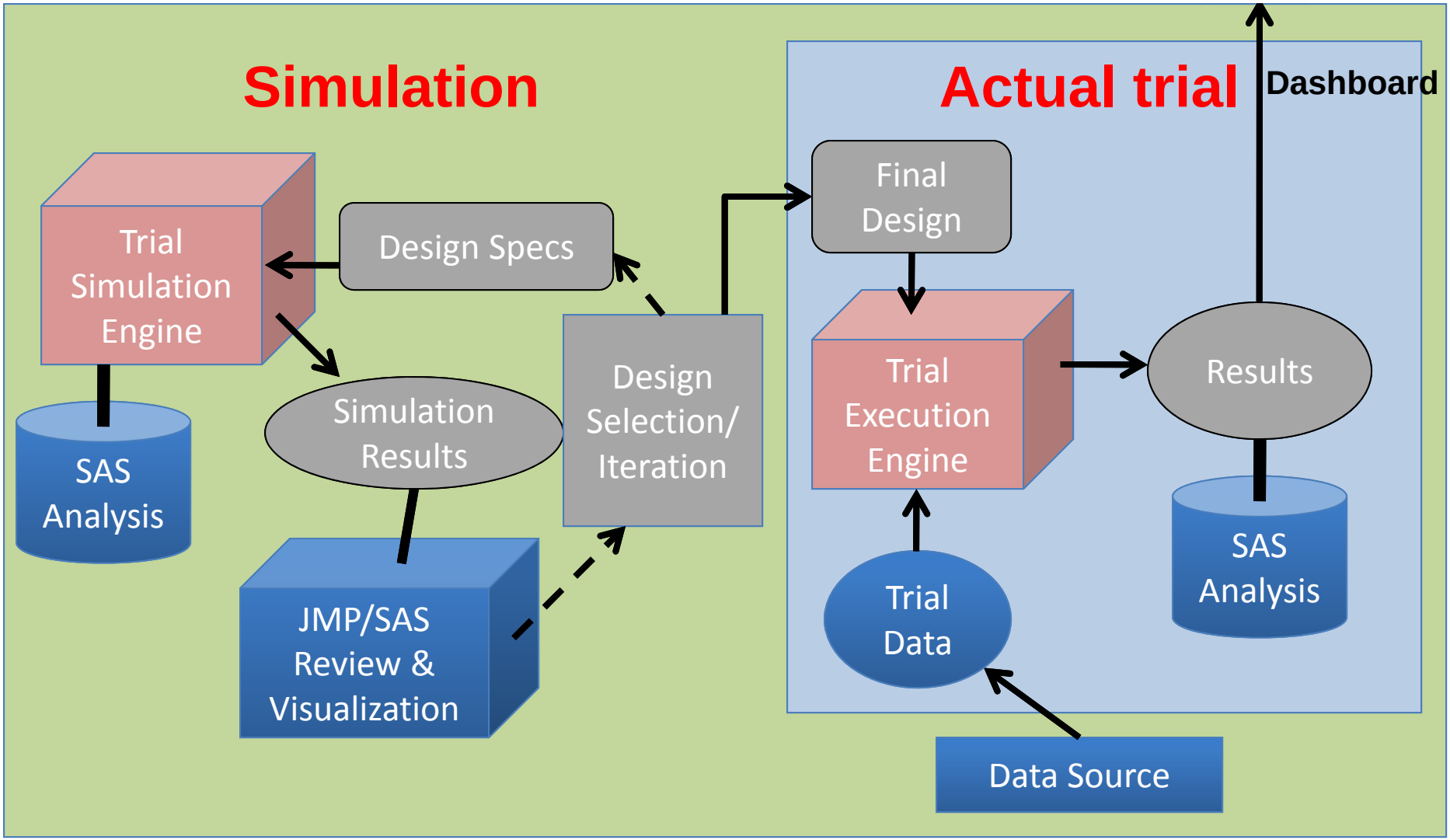
SAS ADAPT?

Berry Consultants
Statistical Innovation

BeST: Solution for Smarter Clinical Trials

Solution Design

DSMB



Also, Selected Short Subject …

Motivated by dinner with Joel Tsevat and John Perentisis …

Designing a clinical trial is making a decision

Decision analysis as a medical research paradigm

- Why do we carry out clinical trials of experimental products?
  - (1) To see if they are safe and effective.
  - (2) To see if the results are unusual assuming product is ineffective.
  - (3) To deliver good medicine to patients.
- Does your answer matter?

Standard Approach to Choosing Sample Size

- Consider time to event. Want
  - 25% reduction in hazards
  - Median for control: 8 months
  - Follow-up after accrual: 12 months
- Answer: n = 650

Where are the questions about the disease? the population? Is this CABG or a rare pediatric cancer?

We need a new paradigm!

Choosing sample size using decision analysis

- Goal: Effective overall treatment of patients, both
  - those who come after the trial and
  - those in the trial
- Example formalizing this goal: Maximize expected number of successes over *all* patients

Compare Joffe/Weeks JNCI Dec 18, 2002

"Many respondents viewed the main societal purpose of clinical trials as benefiting the pa

ge

th **ent**

ar

or **ns**

of patients in trials, conflicts with established principles [from Belmont Report] of research ethics."

We must rethink the "established principles of research ethics"

"More ethical" approach: Maximize overall benefit

- What is "overall"?
- All patients who will be treated with therapies assessed in trial
- Call it N, "patient horizon"
- Enough to know magnitude of N: 100? 1000? 1,000,000?

- Goal: maximize expected number of successes in N
- Either one- or two-armed trial
- Suppose n = 1000 is right for one trial & N = 1,000,000
- Then for other N's use n =

Optimal sample size for one trial and first of two trials					
N	1,000,000	100,000	10,000	1000	100
One trial	1000	320	100	32	10
One/two	170	78	36	17	8

Ratio of sample sizes within row is general

Ratio across rows applies for particular prior distribution

Prior distributions?

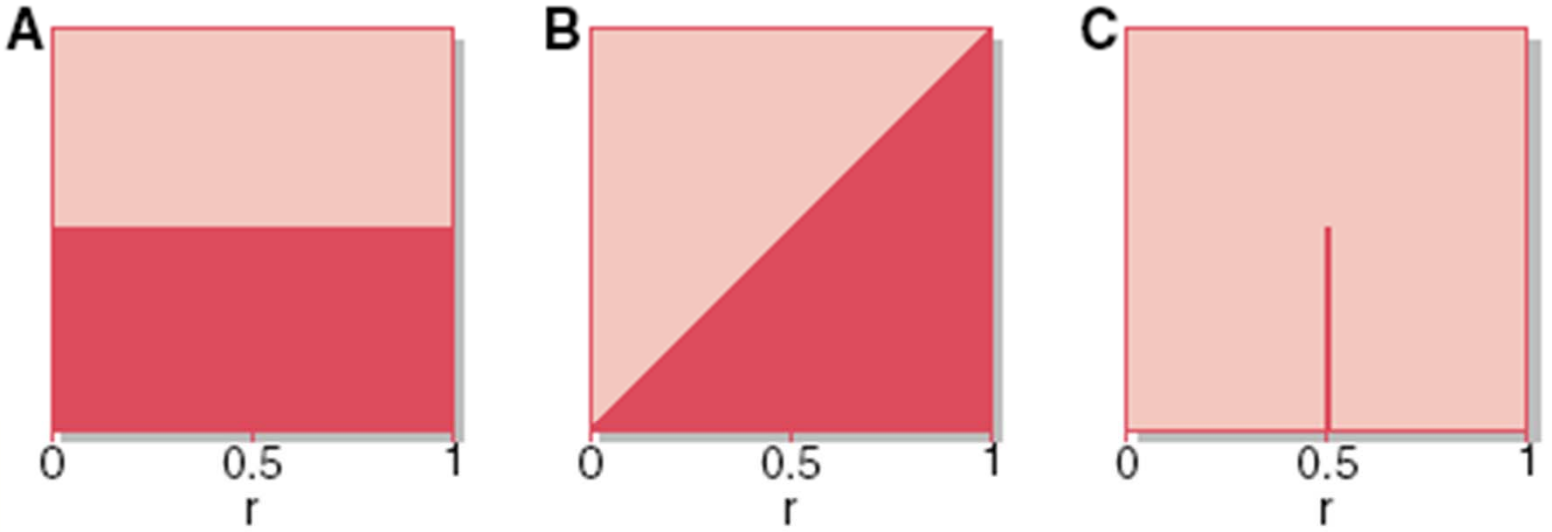
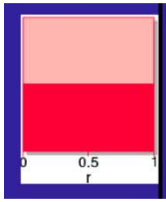
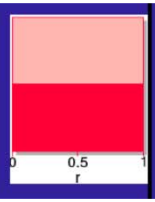
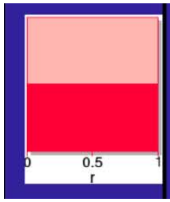
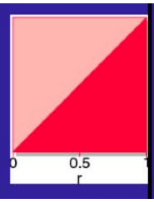
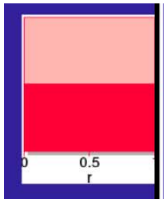
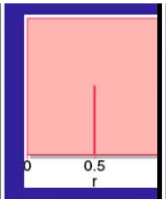


Figure 33-12 Three different prior distributions for r, rate of success. Under distribution A, r is equally likely to be any value between 0 and 1. The density in B is proportional to r, which means, for example, that r greater than 0.5 is three times as probable as r less than 0.5. (These two distributions are the same as the two in Figure 33-1, "a.") Under distribution C, all the probability is concentrated on $r = 0.5$ and so in this case the arm's effectiveness is assumed to be known.

Optimal allocations in a two-armed trial

Optimal allocations to Arms 1 and 2

	Prior distributions								
N			Succ prop			Succ prop			Succ prop
100	6	5	.63	4	8	.71	9	0	.60
1000	21	20	.65	16	30	.74	29	0	.62
10,000	70	69	.66	56	98	.75	99	0	.62
Large N	$\sqrt{\frac{N}{2}}$	$\sqrt{\frac{N}{2}}$	$\frac{2}{3}$	$\sqrt{\frac{N}{3}}$	$\sqrt{N}$	$\frac{3}{4}$	$\sqrt{N}$	0	$\frac{5}{8}$

Conclusions

- Standard clinical research paradigm doesn't apply to rare populations
- Decision analysis addresses right questions, hypothesis testing does not
- Bayesian adaptive approach balances experimentation and treating trial participants effectively
- Bayesian attitude uses all info
- "All info" includes longitudinal info and borrowing adult info to children