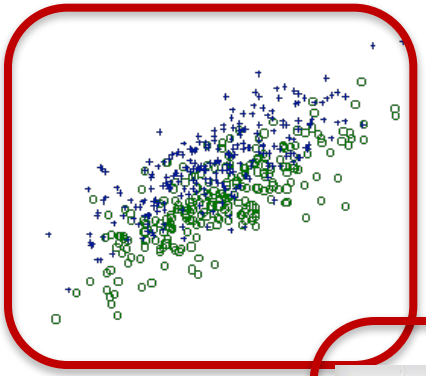
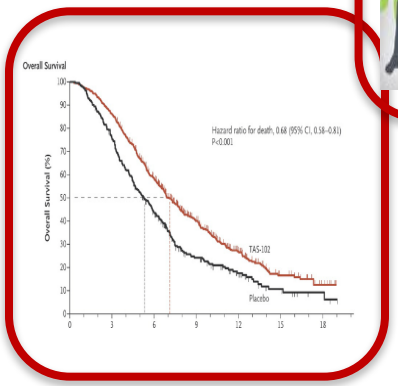
Effect Size and Clinical Population in RCTs





Peter G. Kaufmann Ph.D.

Division of Cardiovascular Sciences
National Heart, Lung and Blood Institute
Bethesda, Maryland

Department of Psychology
American University

International Behavioral Trials Network
Montreal, QC
May 19, 2016

NIH National Heart, Lung, and Blood Institute

Disclosures

- I have no financial conflicts.
- The opinions expressed are my own and do not reflect those of any Agency of the US.

The real purpose of the scientific method is to make sure that Nature hasn't misled you into thinking you know something you actually don't know.

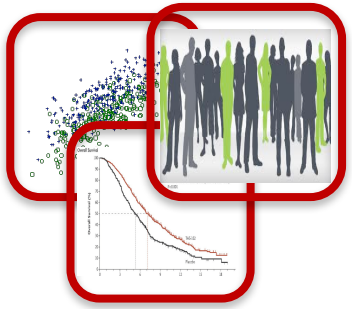
Robert Pirsig
Zen and the Art of Motorcycle Maintenance

Objective

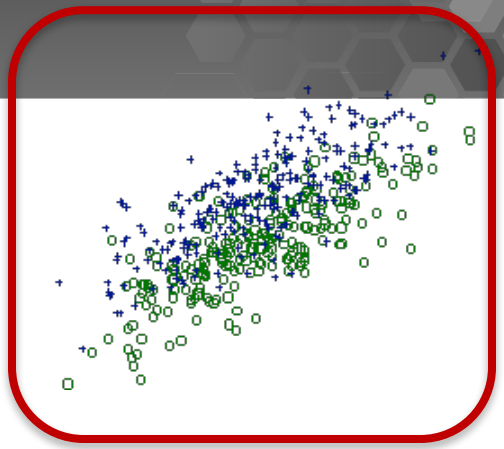
- To illustrate three aspects of clinical trials that, *apart from the effectiveness of the intervention being evaluated*, will determine the outcome.

Overview

- Estimated effect size and “delta inflation.”
- The Minimal Clinically Important Difference (MCID).
- The influence of eligibility criteria on the outcome of a clinical trial.
- The concept of “enrichment” of the study population.



Two Deltas of Interest



- The difference in the treatment delivered to the Intervention and Comparator arms

- The effect size, defined as the primary outcome of the clinical trial.
  - Predicting the effect size is a critical task in clinical trial design.

Parameters in sample size estimation

- The statistical approach:
  - Significance level
  - Power
  - Predicted value of the health status measure in the comparator arm (e.g. event rate)
  - Predicted effect size
- The expedient approach:
  - Stipulate resources and time
  - Estimate number of patients
  - Calculate Delta using conventional “p” and power.

A view of expedience in power calculations

Good lies need a leavening of truth
to make them palatable.

William McIlvanney, 1983

Estimated effect size

Aberegg *et al. Critical Care* 2010, **14**:R77
http://ccforum.com/content/14/2/R77

RESEARCH Open Access

Delta inflation: a bias in the design of randomized controlled trials in critical care medicine

Scott K Aberegg*[1], D Roxanne Richards[2] and James M O'Brien[3]

Abstract

Introduction: Mortality is the most widely accepted outcome measure in randomized controlled trials of therapies for critically ill adults, but most of these trials fail to show a statistically significant mortality benefit. The reasons for this are unknown.

Conclusions: Investigators of therapies for critical illness systematically overestimate treatment effect size (delta) during the design of randomized controlled trials. This bias, which we refer to as "delta inflation", is a potential reason that these trials have a high rate of negative results.

Delta inflation

- Defined as: a biased overestimation of effect size during trial design.

- Consequences: high rate of Type II error, resulting in rejection of potentially valuable interventions.

Method

- Identify critical care RCTs with mortality as the primary outcome in:
  - BMJ
  - NEJM
  - JAMA
  - Lancet
  - Annals of Internal Medicine
  - 1999-2009

Results

- Difference between predicted and observed Delta:

  Only 2/38 trials (5.3%) showed Delta = or > than predicted

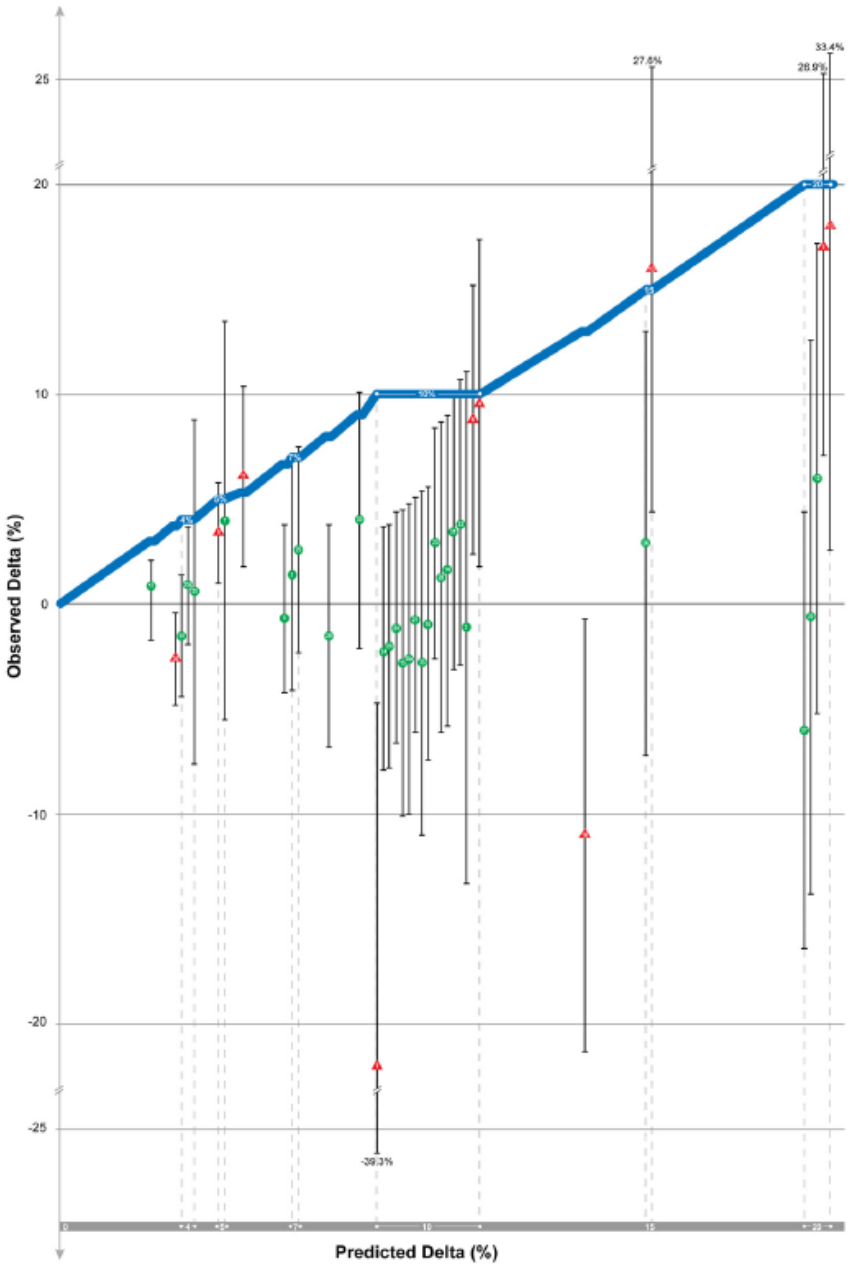
- Mean predicted and observed Delta values for all trials

  (10.1%) (1.4%)

  The difference = Delta-gap =8.7%; ($p<0.0001$)

- 7 of the 38 trials showed unadjusted statistically significant Delta in the hypothesized direction (Red triangles in Fig 1).
- 17 of 38 trials had a negative Delta – treatment worse than the comparator (3 of these terminated early for harm).

Observed vs. Predicted Delta



N=38 trials

- Statistically Non-significant trials
- Statistically Significant trials

NIH National Heart, Lung, and Blood Institute

Aberegg et al Critical Care 2010

Results

- For trials with non-significant effects larger than 3%, the sample size that would be needed for a future trial powered for the observed Delta:
  - Required sample sizes would be 380% to 1,100% larger.

Sample Sizes Under Various Scenarios

	Standard Scenario	Relaxed significance level	Relaxed Power	Baseline Mortality shifted away from 50%	Inflated delta
Significance level (two-sided)	0.05	**0.1**	0.05	0.05	0.05
Power	90%	90%	**80%**	90%	90%
Baseline (placebo) mortality rate	50%	50%	50%	**40%**	50%
Delta (ARR)	10%	10%	10%	10%	**15%**
Required sample size	**1076**	**884**	**816**	**992**	**480**

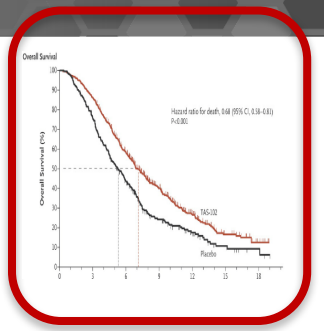
Inflation of delta has a substantially larger impact on required sample size than changes in the other variables. ARR, absolute risk reduction.

Aberegg et al Critical Care 2010

Authors' Conclusions

- Underpowered trials may lead to premature abandonment of promising therapies.
- Underpowered trials may waste resources.
- The use of mortality as the only accepted primary outcome in critical care should be reconsidered.
- Underpowered, delta-inflated trials may be unethical.
  - Participants take risks because they expect the research question to be answered, but usually this can't happen in underpowered trials.

Objective determination of effect size



- Minimal Clinically Important Difference (MCID)
  - Often cited.
  - Can include patients' perspectives.
  - Not an individual decision – should use consensus methods.
  - Can be used to assess primary, secondary or intermediate outcomes.

Adapted from McGlothlin and Lewis JAMA 2014

MCID as intermediate outcome

Enhancing Recovery in Coronary Heart Disease patients (ENRICHD)

Hypothesis: Treating depression post-MI will reduce mortality and re-infarction.

- N= 2,481 patients recruited immediately after MI
- Primary outcome: CV death or reinfarction

Treatment of Depression: CBT plus SSRI as needed

Intermediate outcome: Hamilton Depression Scale

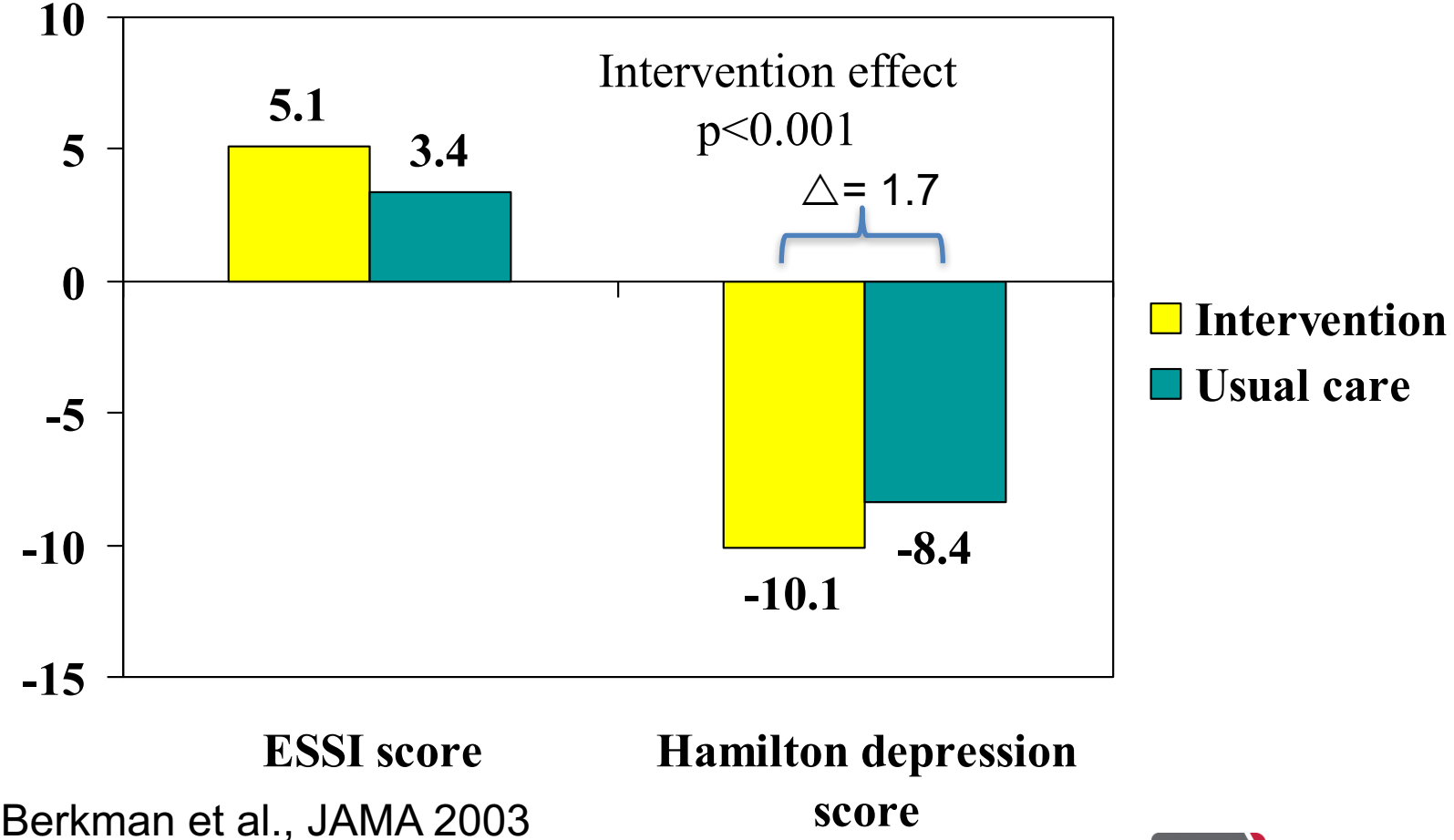
MCID: 2 points in the HamD

- Stipulated in advance by clinical judgment of two therapists involved in the trial.

See Berkman et al., JAMA 2003

ENRICHD Intervention Effect

Baseline to 6-month Changes in Social Support and Depression, ENRICHD



See Berkman et al., JAMA 2003

ESSI reported for patients with low social support only
Hamilton depression score reported for depressed patients only

MCID

- Methods of determining MCID:
  - Statistical
    - Employs statistical distribution characteristics of a validated instrument
  - Consensus
    - Whose perspective is important: clinician, patient?
    - Engage relevant individuals in a consensus process, Delphi method.

Adapted from:
http://www.med.uottawa.ca/courses/CMED6203/Index_notes/Effect Size.htm

Positive and Negative Syndrome Scale (PANSS)

- Data From the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) trial (n = 1,442)
- Clinician-rated CGI-Severity of illness scores 1-7 were compared with PANSS scores of 32.4, 42.2, 57.5, 74.5, 93.0, 110.9, and 131.0, respectively.
- The MCID for PANSS scores = 15.3 points (34.0%) change from baseline.
- A 1.96 SEM on the PANSS corresponded to a 16.5-point (36.2%) change from baseline.
- PANSS for patients in CATIE changed, on average, 14.8%

Hermes et al., J Clin Psychiat. 2012

MCIDs from the Literature

Beck II	17.5% to 32% from baseline depending on D duration Psychol Med. 2015
PHQ-9	5 pts (2 SEMs) on 27 pt. scale Medical Care, 2004
Montgomery-Asberg Depression Rating Scale (MADRS)	1.6 to 1.9 Curr Med Rsch Opinion 2008

Also see:
http://www.med.uottawa.ca/courses/CMED6203/Index_notes/EffectSize.htm

Selecting the Clinical Population

AKA: Eligibility Criteria

Look AHEAD: (Action for Health in Diabetes)
A Critical Analysis

The New England Journal of Medicine

VOLUME 346 FEBRUARY 7, 2002 NUMBER 6

REDUCTION IN THE INCIDENCE OF TYPE 2 DIABETES WITH LIFESTYLE INTERVENTION OR METFORMIN

DIABETES PREVENTION PROGRAM RESEARCH GROUP*

ABSTRACT

Background Type 2 diabetes affects approximately 8 percent of adults in the United States. Some risk factors — elevated plasma glucose concentrations in the fasting state and after an oral glucose load, overweight, and a sedentary lifestyle — are potentially reversible. We hypothesized that modifying these factors with a lifestyle-intervention program or the administration of metformin would prevent or delay the development of diabetes.

TYPE 2 diabetes mellitus, formerly called non-insulin-dependent diabetes mellitus, is a serious, costly disease affecting approximately 8 percent of adults in the United States.[1] Treatment prevents some of its devastating complications[2,3] but does not usually restore normoglycemia or eliminate all the adverse consequences. The diagnosis is often delayed until complications are present.[4] Since current methods of treating diabetes remain inadequate, prevention is preferable. The hy-

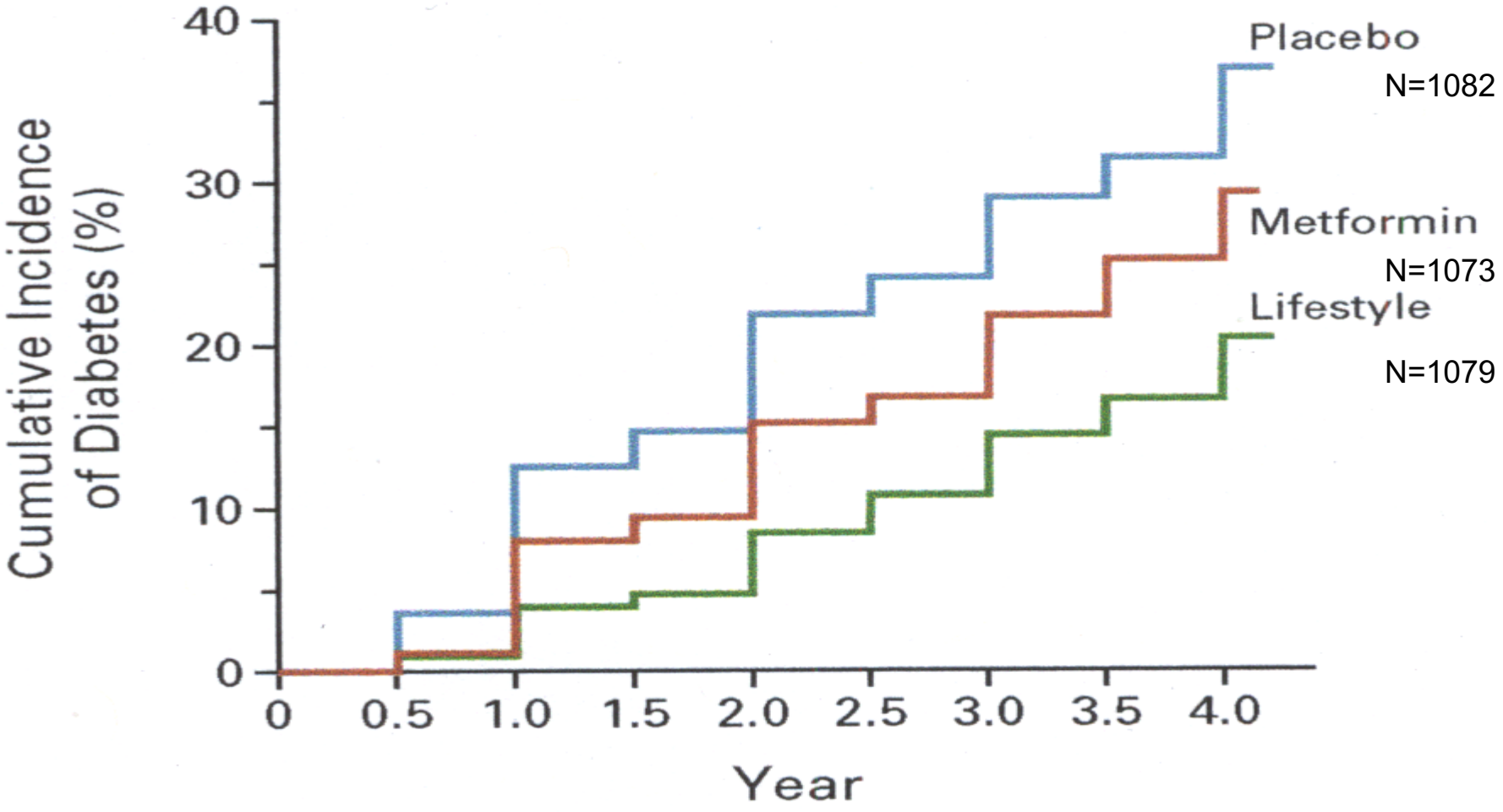


Figure 2. Cumulative Incidence of Diabetes According to Study Group.

The diagnosis of diabetes was based on the criteria of the American Diabetes Association.[11] The incidence of diabetes differed significantly among the three groups ($P<0.001$ for each comparison).

Cardiovascular Effects of Intensive Lifestyle Intervention in Type 2 Diabetes

The Look AHEAD Research Group*

ABSTRACT

BACKGROUND

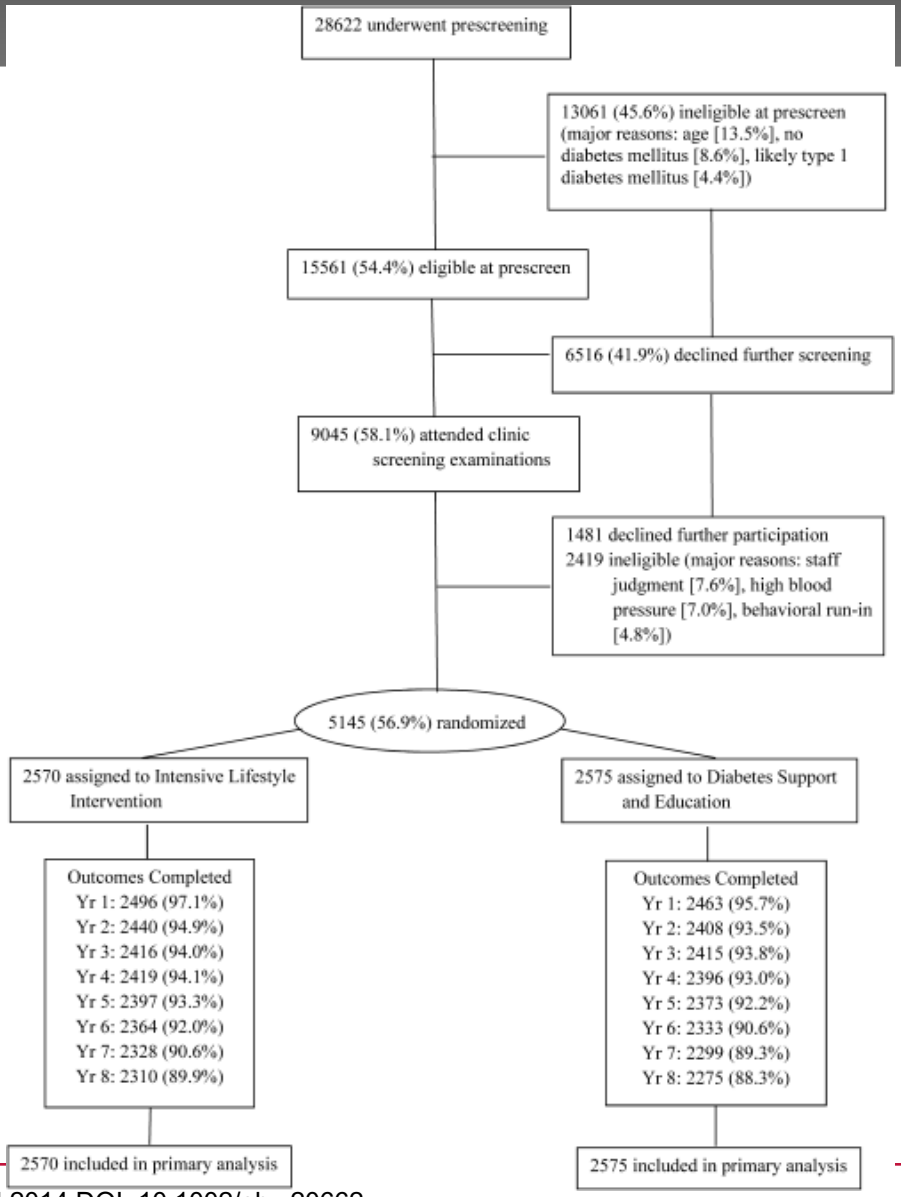
Weight loss is recommended for overweight or obese patients with type 2 diabetes on the basis of short-term studies, but long-term effects on cardiovascular disease remain unknown. We examined whether an intensive lifestyle intervention for weight loss would decrease cardiovascular morbidity and mortality among such patients.

The Look AHEAD Research Group. N Engl J Med 2013;369:145-154.

Eligibility Criteria

- 45-75 years of age
- Type-2 diabetes: self report & verified
- BMI 25+ or 27+ if on insulin
- HbA1c < 11%
- BP < 160/100
- Triglycerides < 600 mg/dl
- Maximal exercise tolerance test, suggesting it was safe to exercise

CONSORT diagram: the Look AHEAD trial.



Obesity
Volume 22, Issue 1, pages 5-13, 11 JAN 2014 DOI: 10.1002/oby.20662
http://onlinelibrary.wiley.com/doi/10.1002/oby.20662/full#oby20662-fig-0001

Baseline Characteristics

Table 1. Characteristics of the Patients at Baseline.*

Variable	Control Group (N=2575)	Intervention Group (N=2570)
Age — yr	58.9±6.9	58.6±6.8
Female sex — no. (%)	1537 (59.7)	1526 (59.4)
Race or ethnic group — no. (%)†		
Black	404 (15.7)	400 (15.6)
Native American	128 (5.0)	130 (5.1)
Asian or Pacific Islander	21 (0.8)	29 (1.1)
White	1631 (63.3)	1621 (63.1)
Hispanic	340 (13.2)	340 (13.2)
Other	51 (2.0)	50 (1.9)

Run-in Period

2-week behavioral run-in period:

-record daily physical activity

-all foods and beverages consumed

-12 of 14 days of record keeping required

Others were ineligible because of their low likelihood of completing the extensive self-monitoring required during treatment.

Intensive Lifestyle Intervention

- Weight loss goal > 7% :
  - Reduced caloric intake
  - Increased physical activity
- First 6 months: weekly sessions (group and individual counseling)
- Toolbox of strategies implemented by staff to help people who had difficulty achieving weight loss,

Diabetes Support and Education

- Three group sessions per year, 4 yrs.
- Education re: diet, exercise, social support
- One group session per year after 4 yrs.

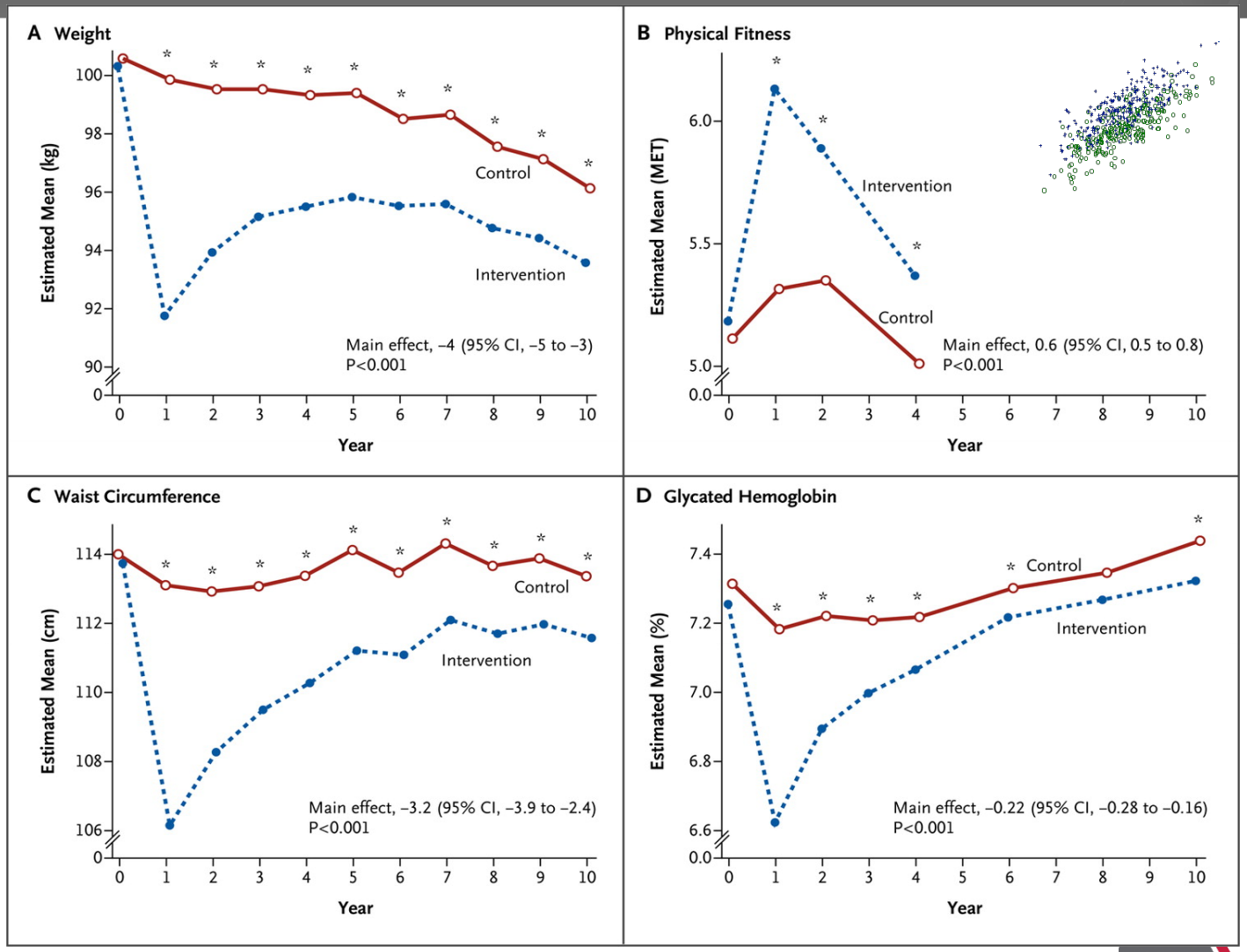
Study End Points

- Primary: First occurrence of
  - CV death
  - Non-fatal MI
  - Non-fatal stroke
  - Maximal FU of 11.5 yrs.

Composite Endpoint

- Hospitalization for angina added after 2 years

Look AHEAD: Changes in Weight, Physical Fitness, Waist Circumference and Glycated Hemoglobin Levels, 10 Years of Follow-up.



NIH National Heart, Lung, and Blood Institute

Trial stopped for futility

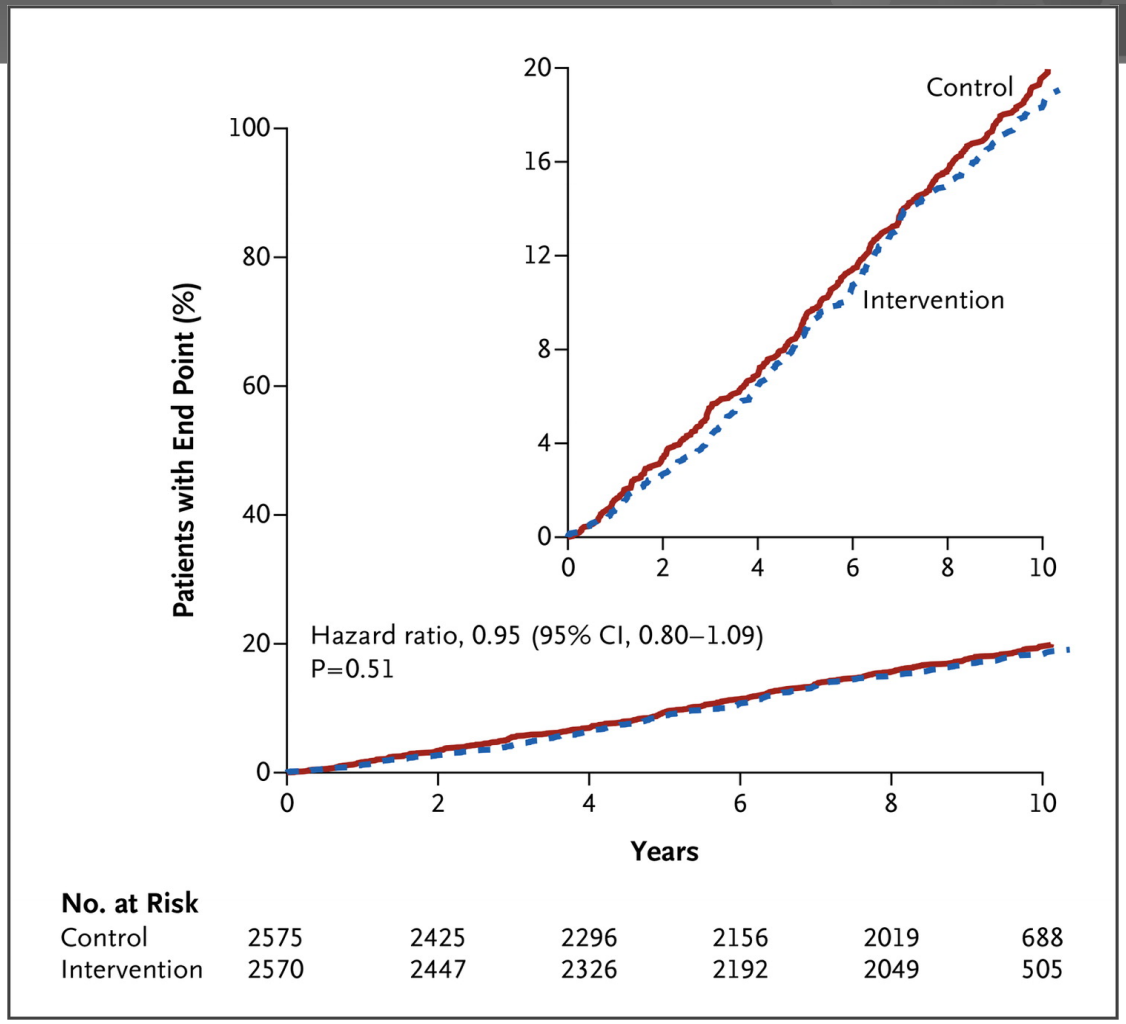
The Good (sort of..):

- Weight loss and physical fitness was as good as it gets.
- When the intervention was stopped September 14, 2012:
  - Median FU was 9.6 years
  - Only 4% were lost to FU

Cumulative Hazard Curves for the Primary Composite End Point.

The Bad:

Primary outcome:
composite of CV death
Nonfatal MI
nonfatal stroke
and
hospitalization for angina



The Ugly:

Originally, Look AHEAD was projected to provide 90% power based on an expected event rate of 3.125% per year in the Diabetes Support and Education group.

A lower-than-expected rate in the first 24 months of follow-up prompted a revision of this expectation: to 80% power based on an event rate of 2.0% per year.

Three years into the trial the actual event rate in the Diabetes Support and Education group was **0.7%** per year.

CLINICAL TRIALS DESIGN

Clinical Trials 2012; **9**: 113–124

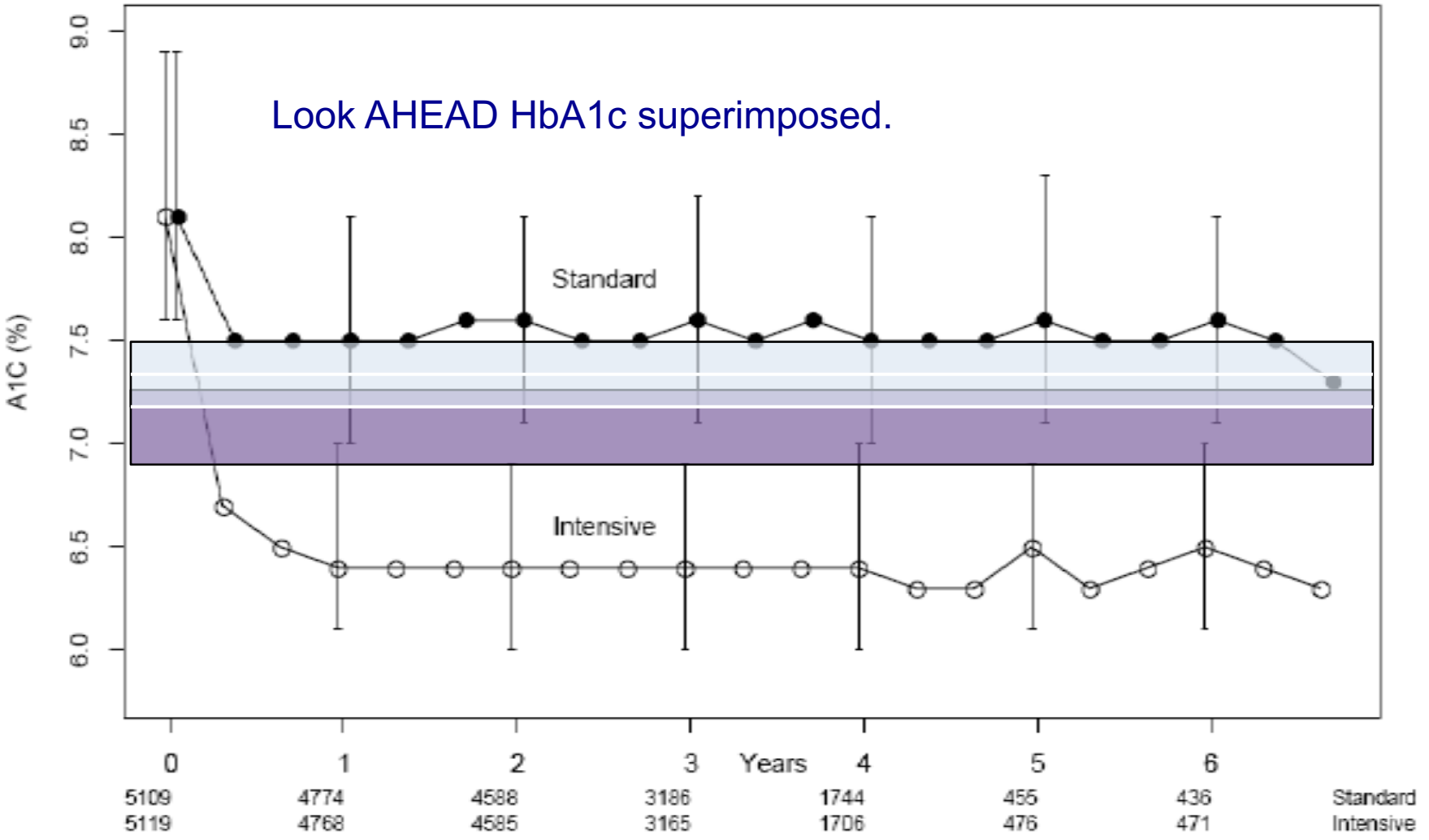
Midcourse correction to a clinical trial when the event rate is underestimated: the Look AHEAD (Action for Health in Diabetes) Study

Frederick L Brancati[a], Mary Evans[b], Curt D Furberg[c], Nancy Geller[d], Steven Haffner[e], Steven E Kahn[f], Peter G Kaufmann[d], Cora E Lewis[g], David M Nathan[h], Bertram Pitt[i] and Monika M Safford[g] on behalf of the Look AHEAD Study Group

The Look AHEAD (Action for Health in Diabetes) Study is a long-term clinical trial that aims to determine the cardiovascular disease (CVD) benefits of an intensive lifestyle intervention (ILI) in obese adults with type 2 diabetes. The study was designed to have 90% statistical power to detect an 18% reduction in the CVD event rate in the ILI Group compared to the Diabetes Support and Education (DSE) Group over 10.5 years of follow-up.

The original power calculations were based on an expected CVD rate of 3.125% per year in the DSE group; however, a much lower-than-expected rate in the first 2 years

ACCORD Median HbA1c



A non-enrichment design

Psychological rehabilitation after myocardial infarction: multicentre randomised controlled trial

D A Jones, R R West

Abstract

***Objective*—To evaluate rehabilitation after myocardial infarction.**

***Design*—Randomised controlled trial of rehabilitation in unselected myocardial infarction patients in six centres, baseline data being collected on admission and by structured interview (of patients and spouses) shortly after discharge and outcome being assessed by structured interview at six months and clinical examination at 12 months.**

***Setting*—Six district general hospitals.**

***Subjects*—All 2328 eligible patients admitted over two years with confirmed myocardial infarction and discharged home within 28 days.**

rehabilitation[18] and therefore could not be cited as an evaluation of psychological therapy.

Published reports implied possible benefit in several different morbidity measures and in cardiac mortality, but even by pooling the findings of all trials the reduction in mortality failed to achieve significance at the 5% level. Furthermore, most trials included only men aged under 65, whereas nearly one third of patients with myocardial infarction are women and nearly half are aged over 65. It was therefore not possible to generalise trial findings to all potentially eligible patients.

Against this background a randomised controlled trial was designed to evaluate rehabilitation by psychological therapy and counselling independent of possible contamination by exercise training or risk factor modifi-

BMJ, 1996

Jones and West intervention – and results

seven two hour outpatient sessions led by clinical psychologists and health visitors.

Principal objectives were (*a*) to give information about the heart and circulation, heart disease, myocardial infarction, treatment and management, and the natural recovery process in order to allay fears and reduce anxiety; (*b*) to increase awareness of stress and stressful situations; (*c*) to teach relaxation skills; (*d*) to improve responses to stressful situations and develop coping skills; (*e*) to promote positive adjustment to illness; and (*f*) to rebuild confidence in patients and spouses. Sessions included teaching, practical exercises with patient participation, group discussion, and individual counselling. The importance of practice

***Results*—At six months there were no significant differences between rehabilitation patients and controls in reported anxiety (prevalence 33%) or depression (19%). Rehabilitation patients reported a lower frequency of angina (median three versus four episodes a week), medication, and physical activity. At 12 months there were no differences in clinical complications, clinical sequelae, or mortality.**

Jones and West - interpretation

- Unselected population – no identified behavioral risk factors
- Weak intervention – anxiety and depression were outcomes, but treatment was non-specific
- Meager results

Summary

- Power calculations should be based on clinically relevant information, not on budgetary considerations.
- Preferably, MCID should be determined on a case-by-case basis.
- Target effect size should be relevant to clinical implications.
- Participants enrolled in clinical trials must have a profile consistent with that of the population on which outcomes are estimated for power calculations.
- The concept of enrichment is an important consideration for detecting a signal that an intervention has a benefit.

Thank you!

Peter G. Kaufmann Ph.D.