

Subgroup Analysis: Lessons Learned from the Montreal Heart Attack Readjustment Trial (M-HART)

Nancy Frasure-Smith, PhD

Associate Professor of Psychiatry, Nursing,
Epidemiology and Statistics, McGill University

and

Senior Research Associate

Montreal Heart Institute Research Center

What Do You Do When Your Well-designed, Carefully Executed Trial Shows No Impact?

- Move on to the next one, that`s why you did the trial
- Less easy to move on in behavioral interventions
 - We have had a relatively small number of adequately powered trials
 - Funding is limited and hard to get (rarely industry-sponsored)
 - Multidisciplinary trials with multiple investigators need multiple publications
 - Despite what we say, behavioral investigators are rarely in a state of equipoise, we want our approach to succeed or, at least, to know why it did not

What Do You Do When Your Well-designed, Carefully Executed Trial Shows No Impact? Ruminare

- Most behavioral interventions are multi-dimensional and evaluated as an overall package
 - Did some aspect work better than others?
- Was the dose big enough? Was it provided for a long enough period of time?
- Was it administered by the right type of health professional?
- Was it administered to the right patients? Not all patients respond well to the same treatment
 - Did it work better for some types of patients?

Subgroup Analysis: You're damned if you do, and you're damned if you don't

- “The scientific challenges and methodologic trap of subgroup analyses” (Furberg & Byington, 1983)
- “..subgroup analysis is both informative and potentially misleading” (Oxman & Guyatt, 1992)
- “..subgroup analyses are important. However, they must be done and interpreted cautiously” (Friedman, Furberg, & De Mets, 1998)
- “The trialist has a duty to analyze by subgroup and to shun data dredging” (Meinert, 1998)
- “Subgroup analyses in clinical trials....fun to look at, but don't believe them” (Sleight, 2000)



Subgroup
Analysis

~~Data
Dredging~~

Definitions (CL Meinert)

- Subgroup analysis: “Assessment of a treatment effect in a subgroup of persons as defined by one or more demographic or entry (baseline) characteristics”
- Data dredging: “Ad hoc subgroup analyses done for the purpose of finding a noteworthy treatment effect as measured by p-value and then presented as ‘proof’ or ‘refutation’ of some hypothesis or contention”

(Meinert, CL. *Teaching slides: Design, Conduct and Analysis of Clinical Trials*, JHU Center for Clinical Trials, Baltimore, 1998)

Definitions (Yusuf, Wittes, Probstfield & Tyroler)

- Proper subgroup: “A group of patients characterized by a common set of ‘baseline’ parameters”
 - Unchangeable characteristics (sex, age)
 - Disease characteristics defined before randomization
- Improper subgroup: “A group of patients characterized by a variable measured after randomization and potentially affected by treatment”
 - Responders vs. non-responders

(Yusuf et al. Analysis and interpretation of treatment effects in subgroups of patients in randomized controlled trials. *JAMA* 1991; 266:93-98,

Why Subgroup Analysis is Needed (The Unpredictability Paradox)

- Patients in trials are always clinically heterogeneous, but trial results apply only to the average patient who took part in the trial
- There is no way for a clinician to know how an individual patient will respond to a “proven” therapy; information about subgroups is more relevant to patient treatment



our **ACTIVITIES**

- about the **Society**
- our **Activities**
- about your **Health**
- understanding **Research**
- funding **Research**
- advocating **Policy**
- making a **Contribution**
- for the **Press**

to discuss issues that impact the recruitment and retention of women in clinical trials and to share successful strategies creating an atmosphere of trust and respect. [To order a copy of the meeting report, click here.](#)

- **Subgroup Analysis and Statistical Design For Detecting Sex Differences in Clinical Trials**

The second workshop, *Subgroup Analysis and Statistical Design For Detecting Sex Differences in Clinical Trials*, held in July 2000, was aimed at statisticians and principle investigators involved in study design and data analysis. The meeting brought together experts to discuss the importance of subgroup analysis, a statistical procedure that takes data from a general group of study subjects and looks for differences within a subset of those subjects that share a specific characteristic, such as sex, age or state of disease. [Click here for a summary of the meeting.](#)

SOCIETY STUDY REVEALS SEX ANALYSIS OF HEALTH DATA LACKING IN FEDERALLY FUNDED STUDIES

Why Subgroup Analysis is Problematic (The Unpredictability Paradox)

- Subgroup analysis permits evaluation of consistency of treatment effects
- When analysis reveals lack of consistency, apparent differences are not often replicated
 - Multiplicity of comparisons (false positives)
 - Inadequate sample sizes in subgroups (false negatives)
 - Lack of stratified randomization in subgroups

Contradictory Results from Different Centers in the Beta-blocker Heart Attack Trial (31 Centers; 11 favored placebo) (Mattocks & Horowitz, Biol Psychiatry, 2000)

Center	Number of patients	Mortality Rate (%)	
		Propranolol	Placebo
1	97	0	6.3
2	118	8.5	1.7
3	115	1.8	12.1
4	127	12.5	4.8
5	113	1.8	8.9
6	110	12.7	7.3
7	84	2.4	9.5
8	150	10.7	6.7
9	193	7.4	6.1

Reproducibility of Ad Hoc Subgroup Comparisons in Trials of Beta-blockers for Acute MI (Yusuf et al, 1991)

Subgroup Benefit	Prior Ho?	Confirmed?	Overall P
Heart rate >100 bpm	No	No	NS
High CV risk	No	No	NS
<65 years	?	No	<0.03
Heart rate > 65 bpm	No	No	NS
Type of ECG changes	No	Inconsistent	NS
Anterior MI	No	No	<0.06
Within 6 months of MI	No	No	No
Type of ECG changes	No	Inconsistent	<0.003
Type of BB	No	Inconsistent	<0.0001

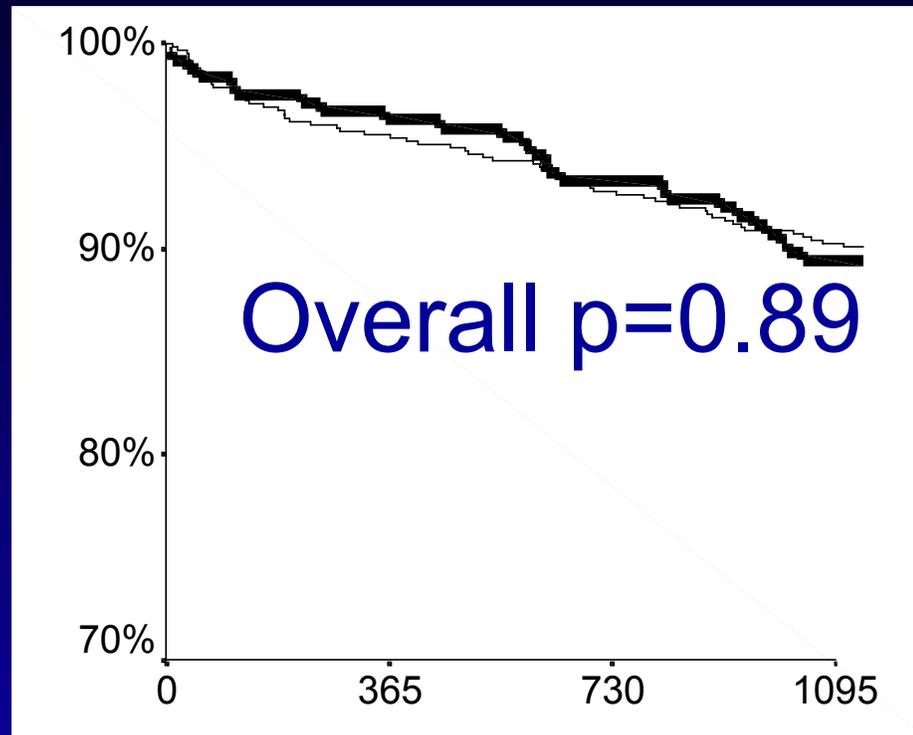
Unreliability of “data-dependant” subgroup analyses:
 ISIS-2 trial of aspirin among over 17,000 patients with
 suspected myocardial infarction (Peto et al, 1988)

	Vascular Death by 1 Month		P
	Aspirin	Placebo	
Astrological birth sign			
Libra or Gemini	150 (11.1%)	147 (10.2%)	0.5
All other signs	654 (9.0%)	869 (12.1%)	<0.001
Any birth sign	804 (9.4%)	1016 (11.8%)	<0.001

THE DATA MINER

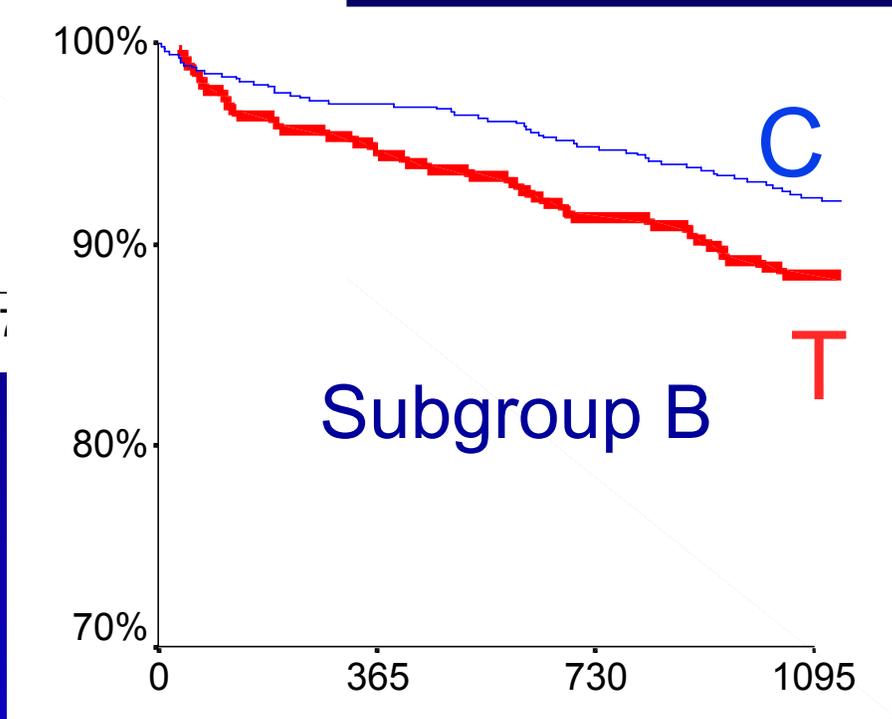
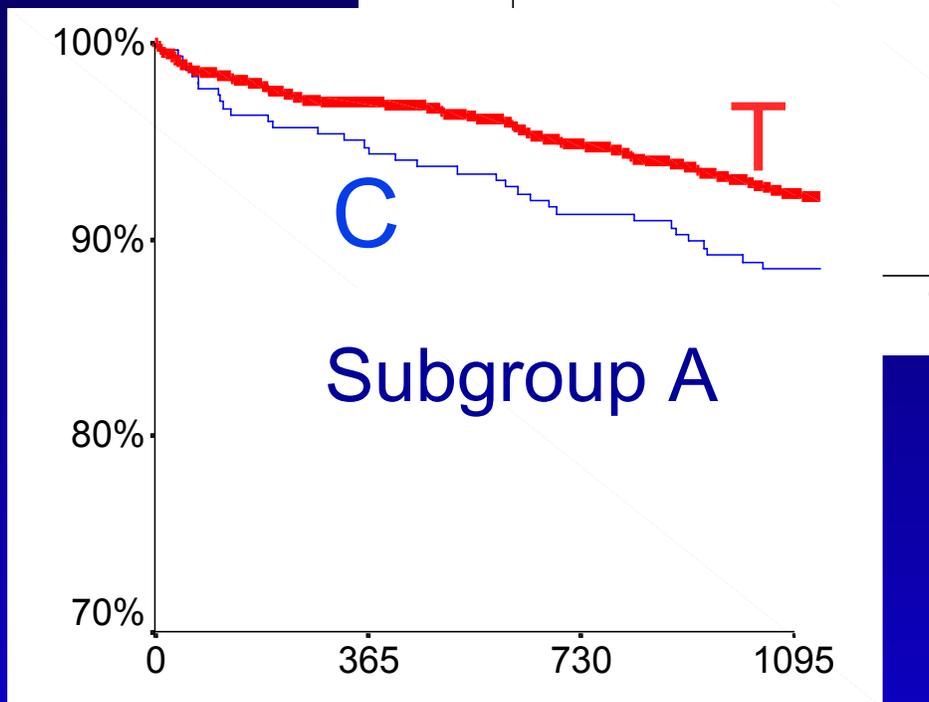
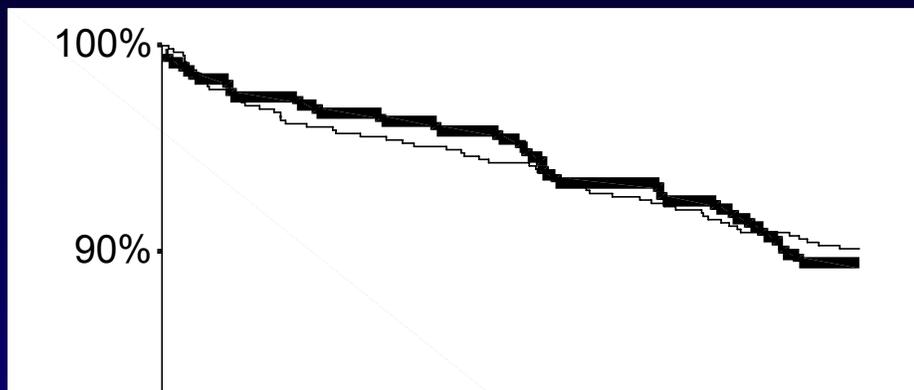
Eureka! I found something significant at $p < 0.05$!





When there is no overall difference between groups, you can almost always find some subgroups with opposite impacts that cancel each other out in the overall analysis

$$0 = +1 + (-1)$$



Low Education	< 65	\geq 65	P-value
Men	19.7%	37.1%	<0.001
Women	34.3%	41.3%	0.23

Lives Alone	< 65	\geq 65	P-value
Men	11.8%	13.4%	0.63
Women	22.9%	42.7%	<0.001

Previous MI	< 65	\geq 65	P-value
Men	25.4%	26.8%	0.75
Women	13.6%	22.8%	0.028

Problem of multiplicity

Which subgroup pairs differ from each other?

Low Education	< 65	≥ 65	P-value	Interaction P
Men	19.7%	37.1%	<0.001	0.18
Women	34.3%	41.3%	0.23	

Lives Alone	< 65	≥ 65	P-value	Interaction P
Men	11.8%	13.4%	0.63	0.040
Women	22.9%	42.7%	<0.001	

Previous MI	< 65	≥ 65	P-value	Interaction P
Men	25.4%	26.8%	0.75	0.10
Women	13.6%	22.8%	0.028	

Bonferoni correction

$$0.05/3 = 0.017$$

Which subgroup pairs differ from each other?

Comparison of Subgroup Analysis and Data Dredging (based on Meinart, 1998)

■ Similarities

- Ad hoc (unless subgroups planned in advance)
- Same analytic approaches
- Based on baseline characteristics

■ Differences

Subgroup Analysis	Data Dredging
Aim: explanation	Aim: proclamation
Cynical about p-values as indicators of truth	P-value fixated
Reluctant to draw conclusions	Predisposed to conclude

How to be a Data Dredger (straw man modified from Meinert, 1998)

- Do an enormous number of subgroup comparisons, ignoring the size of the data set, and never reporting the number of comparisons
- Focus on all p-values <0.05 and ignore the rest
- Choose cut-points for subgroups that maximize differences
- Combine two or more baseline characteristics to create subgroups, if it increases the differences
- Try to sell the results as having major medical implications
- “Stay near the phone awaiting a call regarding your nomination for the Nobel Prize in Medicine, promoting your candidacy for the prize while waiting”



Subgroup
Analysis

~~Data
Dredging~~

Recommendations for Subgroup Analyses

■ Planning phase:

- Pre-specify a small number of biologically appropriate subgroups
- Try to assure that the sample size allows adequate power to test intervention in these subgroups
- Consider stratified randomization in most important subgroups

Recommendations for Subgroup Analyses

■ Analysis phase:

- Ad hoc subgroups more appropriate following an overall positive (or negative) trial, than for a trial with no impact; remember $0 = +1 + (-1)$
- Guide selection of subgroups by biological plausibility
- Do not test all possible subgroups
- Make adjustments for number of comparisons
- Examine interaction effects rather than separate p-values for each member of a pair of subgroups
- Focus on overall pattern of results rather than individual p-values
- Interpret results in light of results of other studies

What Do You Do When Your Carefully Planned and Executed Behavioral Trial is Negative or Shows No Impact?

- A) Bury it, its probably a fluke, the results could be bad for the field
- B) Carryout pre-specified analyses, and publish as soon as possible
- C) Carry out as many subgroup analyses as it takes, until you find something positive
- D) Carry out well-thought out, ad hoc analyses to possibly find out why it did not work, to help in designing new trials

Montreal Heart Attack Readjustment Trial (M-HART)

THE LANCET

Lancet 1997; **350**: 473–79

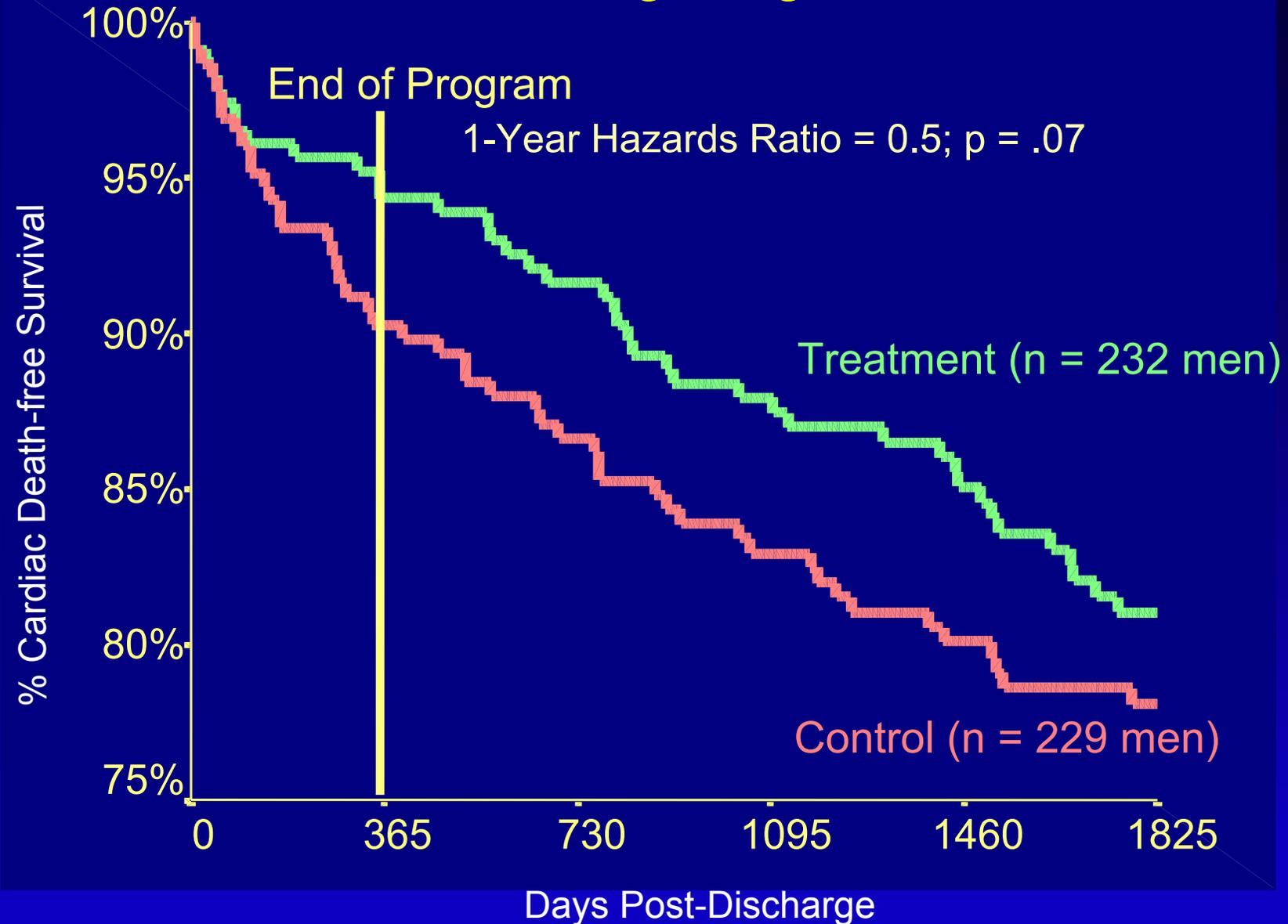
Randomised trial of home-based psychosocial nursing intervention for patients recovering from myocardial infarction

Nancy Frasure-Smith, François Lespérance, Raymond H Prince, Pierre Verrier, Rachel A Garber, Martin Juneau, Christina Wolfson, Martial G Bourassa

Prior to 1990 Only 2 “Large-Scale” Studies of Treatments to Improve Cardiac Prognosis by Altering Psychosocial Factors

- The RCPP (Recurrent Coronary Prevention Project)
 - Recruitment 1977 - 1978
 - Target = Type A behavior
- The IHD Life Stress Monitoring Program
 - Recruitment 1977 - 1981
 - Target = Psychological distress
- Neither included women
- Both had problematic designs because of the way randomization was handled

5-Year Cardiac Mortality in the IHD Life Stress Monitoring Program



Limitations of IHD Life Stress Monitoring Program

- Methodological problems with randomization resulted in imbalances in social class and disease severity; treatment group was at less risk at baseline
- Sample entirely male
- Little evidence of potential mechanisms for program impact
 - Psychological improvement?
 - Compliance improvement?
 - Better medical care?

Montreal Heart Attack Readjustment Trial (M-HART)

- 1-year RCT of usual care vs M-HART Program
 - Funding from NHLBI and Canadian NHRDP
- Program Target: reduction of *psychological distress*
- Primary Objective: To test the hypothesis that program participation would reduce 1-year cardiac mortality in women, as well as men
- Primary Outcome: 1-year cardiac mortality (distal outcome)

Montreal Heart Attack Readjustment Trial (M-HART)

■ Secondary Objective: To collect data on potential behavioral/ psychological mechanisms for program impact (proximal outcomes)

- Changes in negative emotions
- Medication compliance
- Risk factor modification
- Health system contacts

M-HART Program

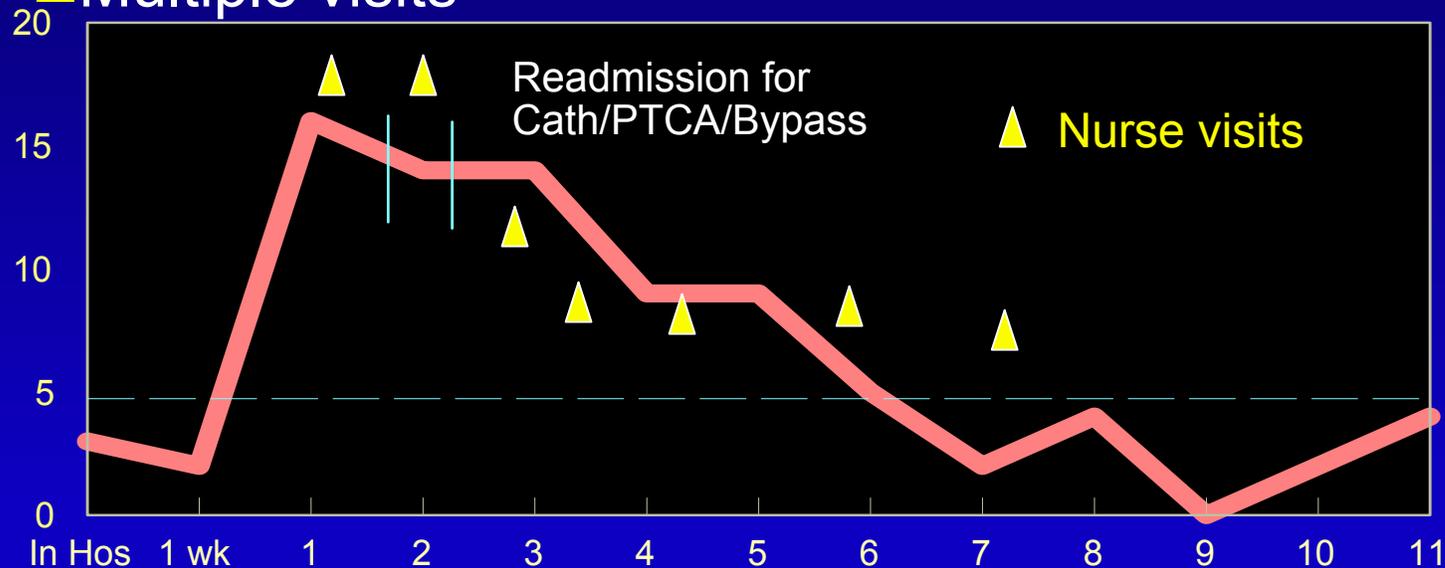
- Monthly phone monitoring of psychological distress (the GHQ-20)

- Intervention for high distress (83% of treatment patients):

 - Cardiac nurses

 - Home-based, case management approach

 - Multiple visits



M-HART Study Design



M-HART Sample Size Calculation

- 2-sided alpha of 0.05
- Power of .80
- Expected 1-year cardiac mortality rates based on existing literature in 1990
 - 10% in men, 12% in women
- Ability to detect a reduction of 50% or more in mortality
 - 896 men
 - 734 women

M-HART Recruitment (1/91 – 9/94)

- 6415 admissions to 10 Montreal area hospitals
- 4047 met research criteria for myocardial infarction
 - 2 of 3 criteria: typical chest pain, 2 X upper limit of normal for CPK and new Q-wave on ECG
- 2483 met other study criteria
 - no other life threatening condition, able to complete baseline interview in hospital, spoke French or English, had access to telephone, lived within 20 miles of study hospital
- 2180 approached for participation (303 discharged early)
- 1376 accepted, completed interview and were randomized
 - 69% of men (903)
 - 54% of women (473)

■ After 2 years of recruitment

- 903 men randomized (target 896)

■ After 3 years of recruitment

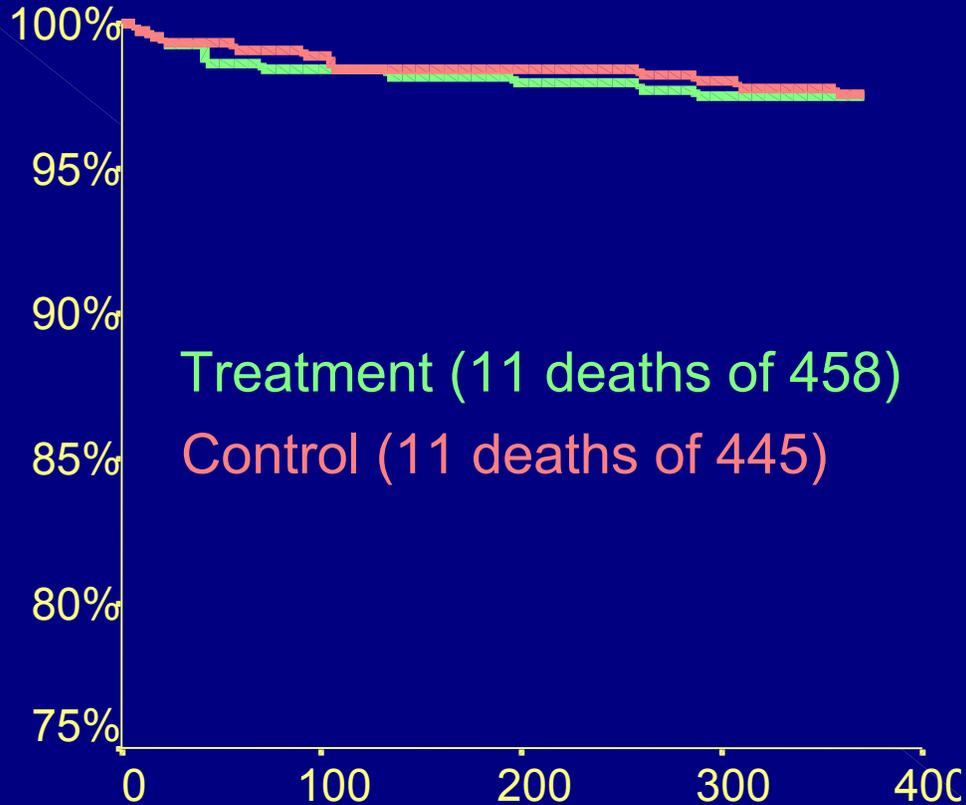
- 473 women randomized (2/3 of target 734)
- Target required an additional 20-months of recruitment in current hospitals
- What to do?

1 Year Cardiac Mortality

Cumulative
Survival

Men

Women



Hazards Ratio = 0.97; $p = .94$
(95% CI: 0.42 - 2.23)

Hazards Ratio = 1.92; $p = .064$
(95% CI: 0.95 - 3.88)



“Four years of research, and now you tell me you forgot which is the control group!”

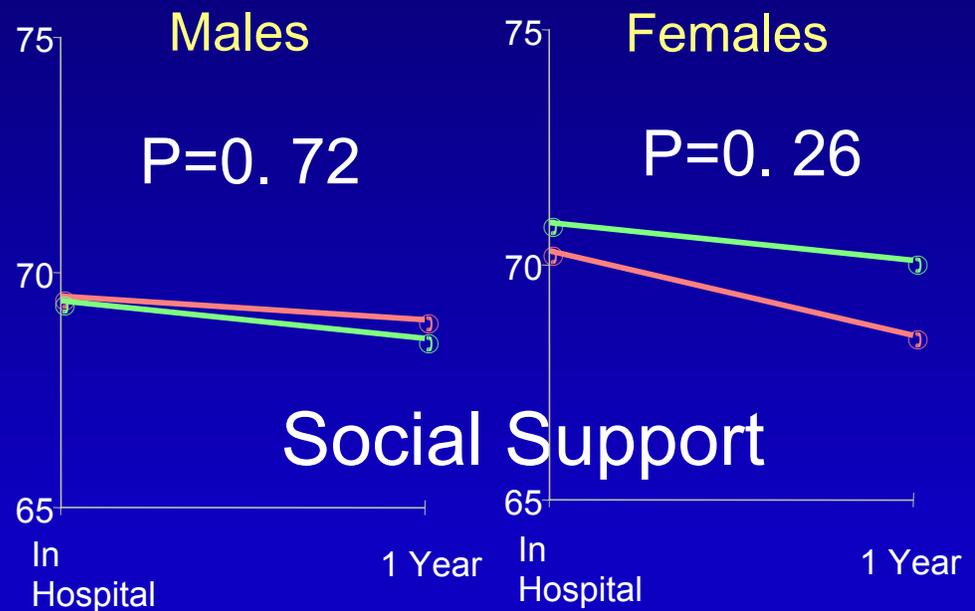
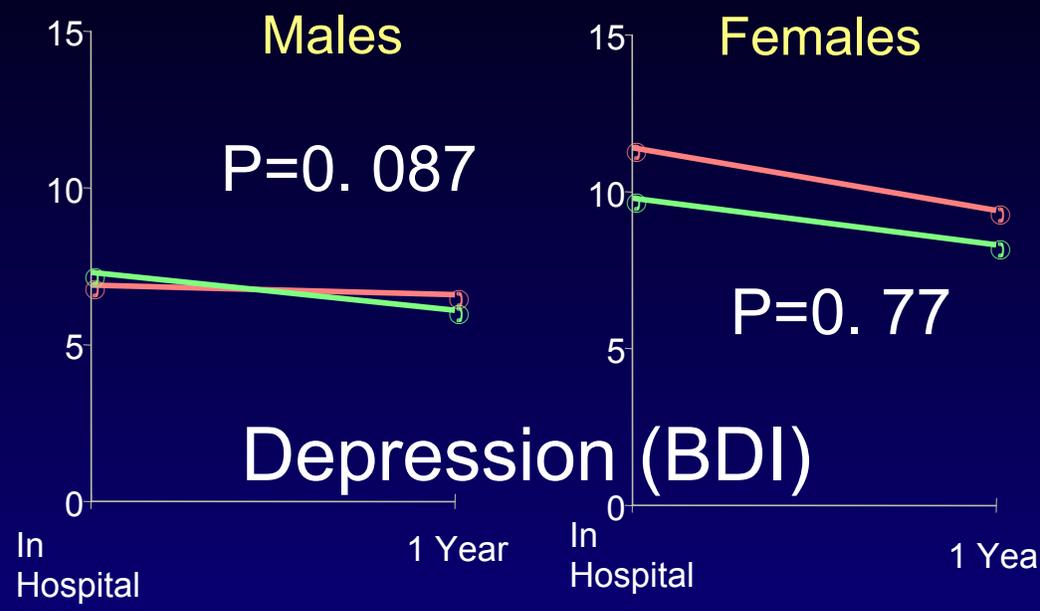
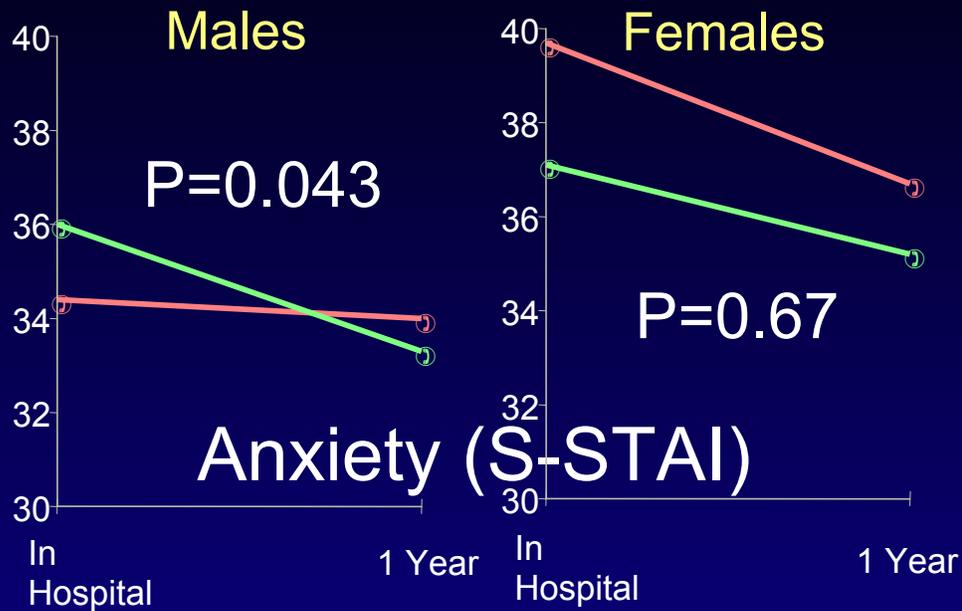
G. Spitzer, APA Monitor, August 1971.

Power Calculations for M-HART Men

- Target sample based on 10% cardiac mortality with usual care; observed rate was 2.5%
- Achieved Power ($\alpha=.05$; 2-tailed) = 28%
- To detect a 50% difference in mortality would have required a total of 3020 male patients (1510 per group)
- To detect a 25% difference would have required 15,400 male patients (7700 per group)
- One of the biggest psychosocial intervention trials for post-MI patients was not big enough

Pre-specified Analyses in M-HART

- Mechanisms to explain beneficial impact
 - Medication compliance
 - Risk factor changes
 - Health care system usage
 - Negative emotions
- Nothing was planned to explain lack of impact or negative impact; despite the 2-sided power calculation, we never imagined it was possible



Possible explanations for outcome in women

- Imbalances in baseline characteristics? No
- Program reduced MD contacts? No
- Program influenced medication prescription or compliance? No
- Program influenced risk factor modification? No
- Program Impact not sufficient on negative emotions?
 - Cut-point for intervention too low (81% of men, 86% of women): Interventions for minor levels of distress
 - Monthly phoning may have interfered with denial
 - Not all patients cope well with increased information

Subgroup Analyses in M-HART

- 1) Ad hoc, systematic analysis of all two-way interactions between baseline factors and treatment/control group
- 2) Ad hoc analysis of outcomes for treatment responders vs. non-responders
- 3) Long-term ad hoc analysis to explore the possibility that the program may have interfered with normal coping in some patients, paradoxically increasing distress

Ad Hoc Subgroup Analyses

1) Systematic Assessment of All Two-way Interactions of Group (Treatment/Control) by 17 Baseline Variables



6 (35%) nominally significant or near significant Interactions (2 by chance)

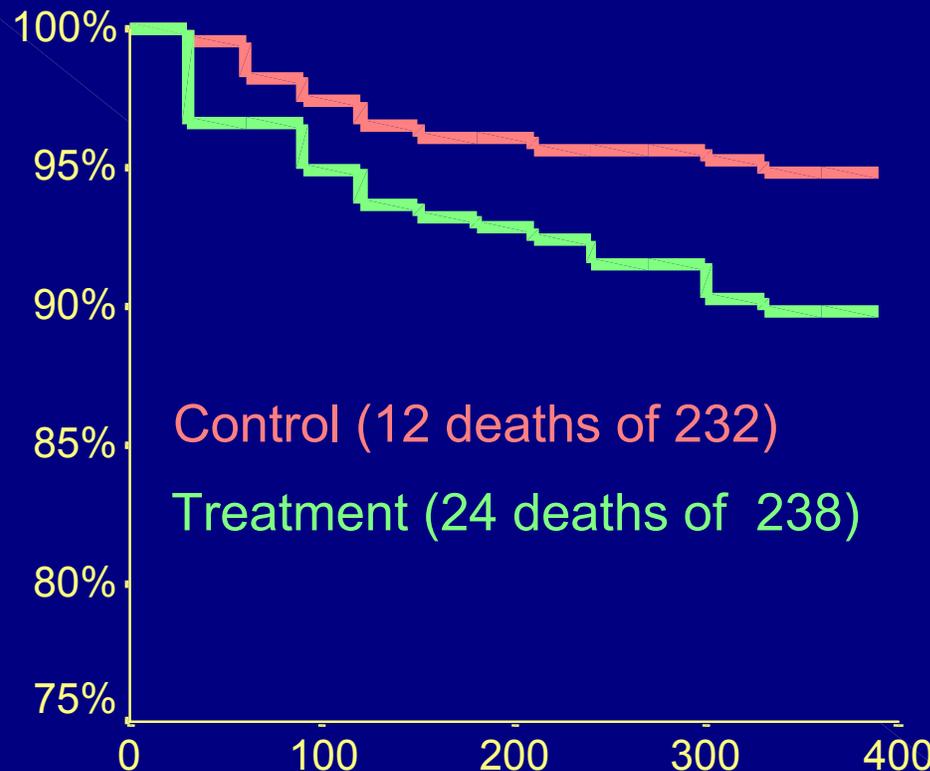
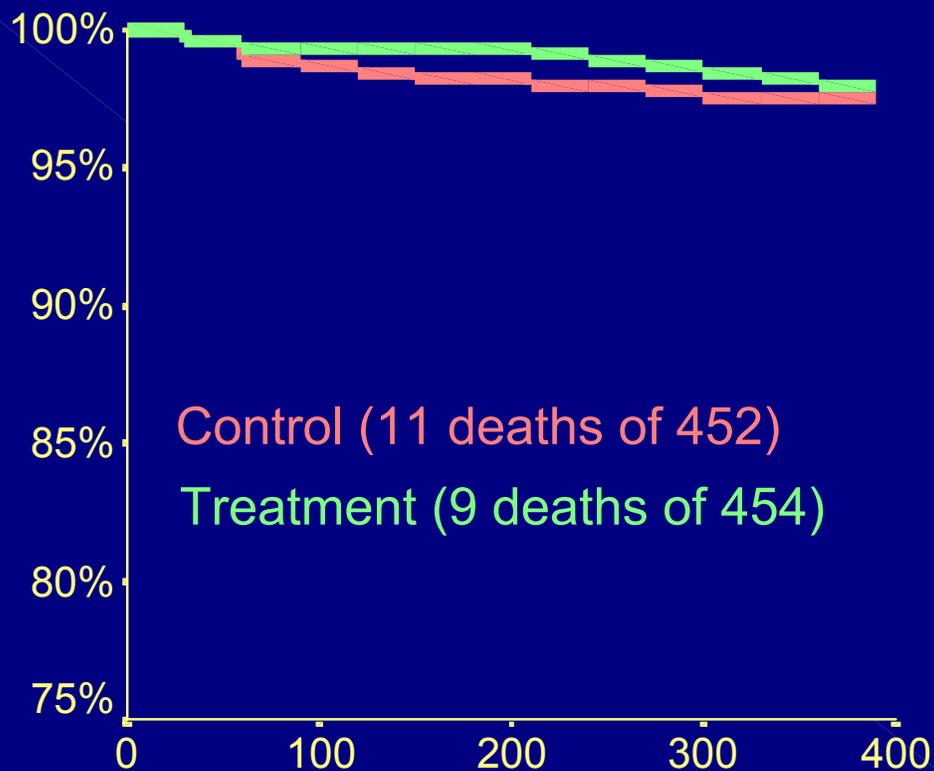
Age by Group	p=0.10
Education by Group	p=0.027
Q Wave MI by Group	p=0.012
Anger Expression by Group	p=0.069
No Close Friends by Group	p=0.066
Living Alone by Group	p=0.016

1 Year Cardiac Mortality

Cumulative Survival

< 65 Years

> or = 65 Years



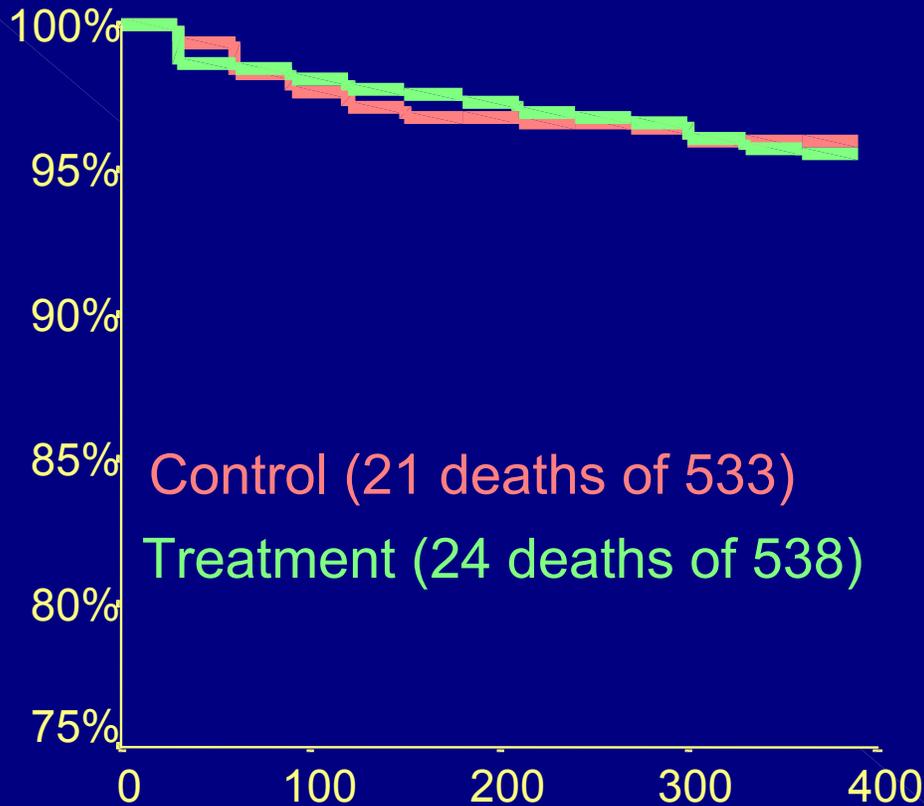
Days Post-Discharge

Hazards Ratio = 0.81; p = .63
(95% CI: 0.33 - 1.95)

Hazards Ratio = 2.01; p = .048
(95% CI: 1.01 - 4.03)

1 Year Cardiac Mortality

Cumulative Survival
At Least 1 Close Friend

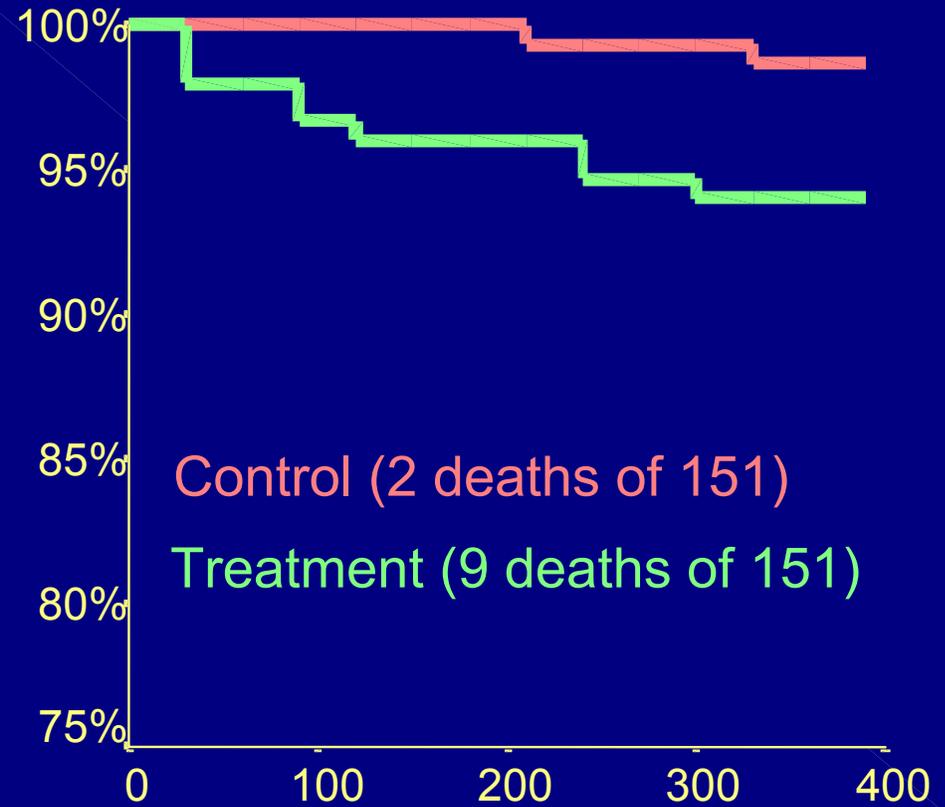


Control (21 deaths of 533)

Treatment (24 deaths of 538)

Hazards Ratio = 1.13; $p = .68$
(95% CI: 0.63 - 2.03)

No Close Friends



Control (2 deaths of 151)

Treatment (9 deaths of 151)

Hazards Ratio = 4.62; $p = .050$
(95% CI: 1.00 - 21.38)

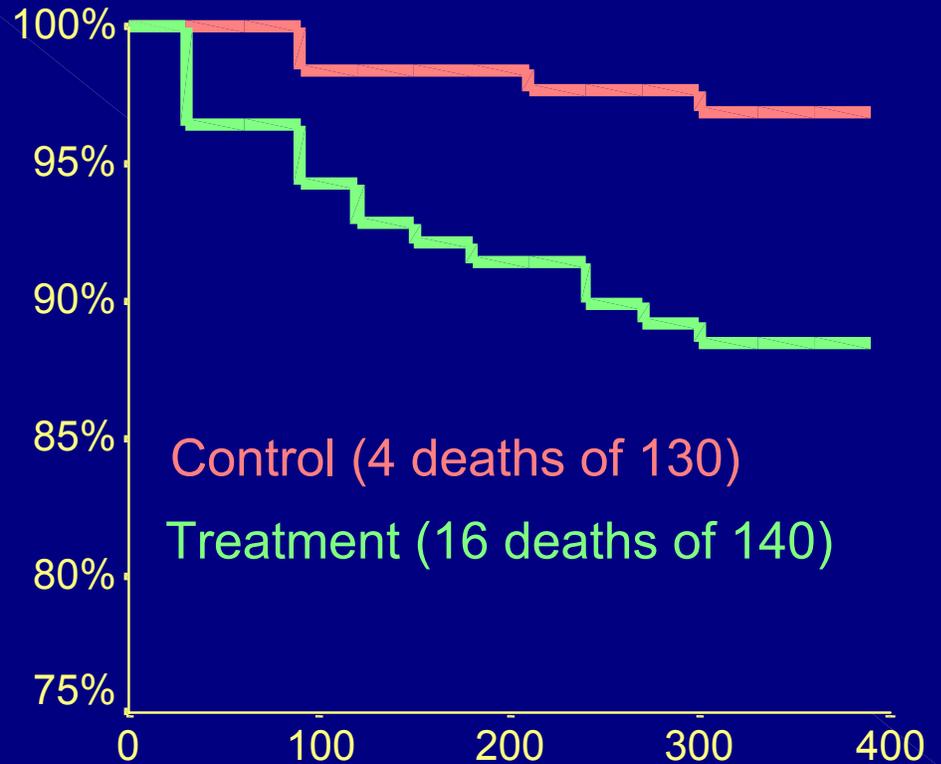
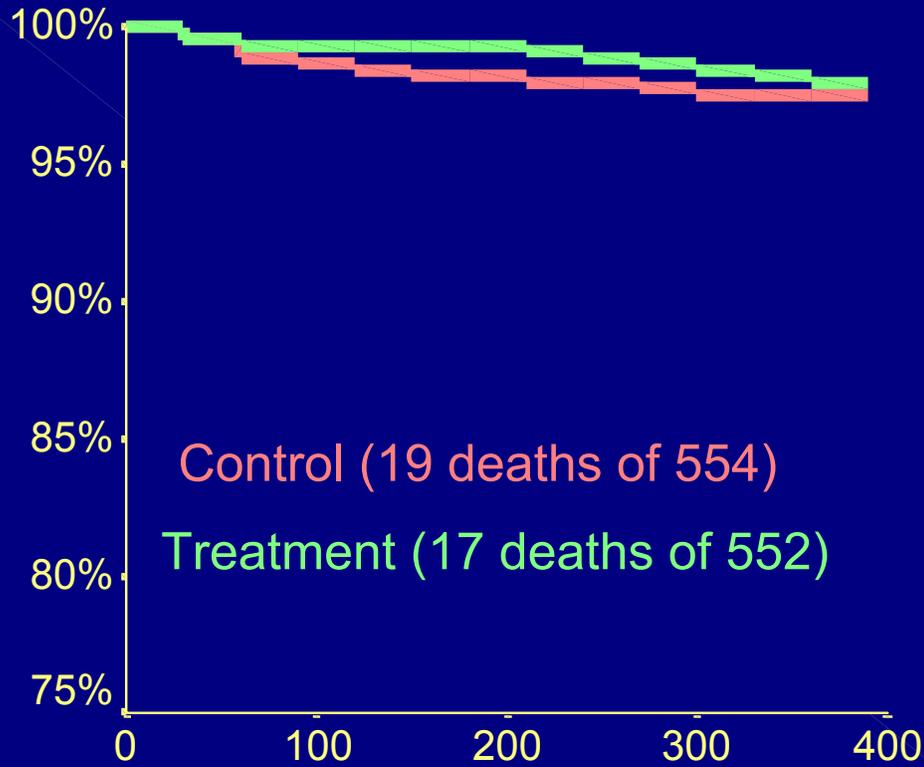
Days Post-Discharge

1 Year Cardiac Mortality

Cumulative Survival

Not Living Alone

Living Alone



Days Post-Discharge

Hazards Ratio = 0.89; $p = .73$
(95% CI: 0.46 - 1.72)

Hazards Ratio = 3.93; $p = .014$
(95% CI: 1.31 - 11.75)

Summary of Systematic Ad Hoc Subgroup Analyses

- Several subgroups in addition to women were at somewhat increased risk with the treatment
 - 65 years and older
 - Q-wave infarctions
 - Low anger expression
 - No close friends
 - Living alone
 - Low education
- No evidence of any subgroups with a positive benefit
- Differences for patients with no close friends and those living alone were particularly disturbing
- After conference presentation, never tried to publish results because of controversies over subgroup analyses

Acceptable or not?

1) Ad hoc, systematic analysis of all two-way interactions between baseline factors and treatment/control group

- Proper subgroups (+)
- Interaction approach more appropriate than multiple individual comparisons (+)
- Multiplicity (false positives) (-)
- Small sample sizes (false negatives) (-)

Subgroup Analyses in M-HART

- 2) Ad hoc analysis of outcomes for treatment responders vs. non-responders
 - Perhaps program failed because the treatment did not work in enough patients.
 - What types of patients showed an improvement in psychological distress in response to nursing visits?
 - Did improvements in psychological distress translate into improvements in prognosis?

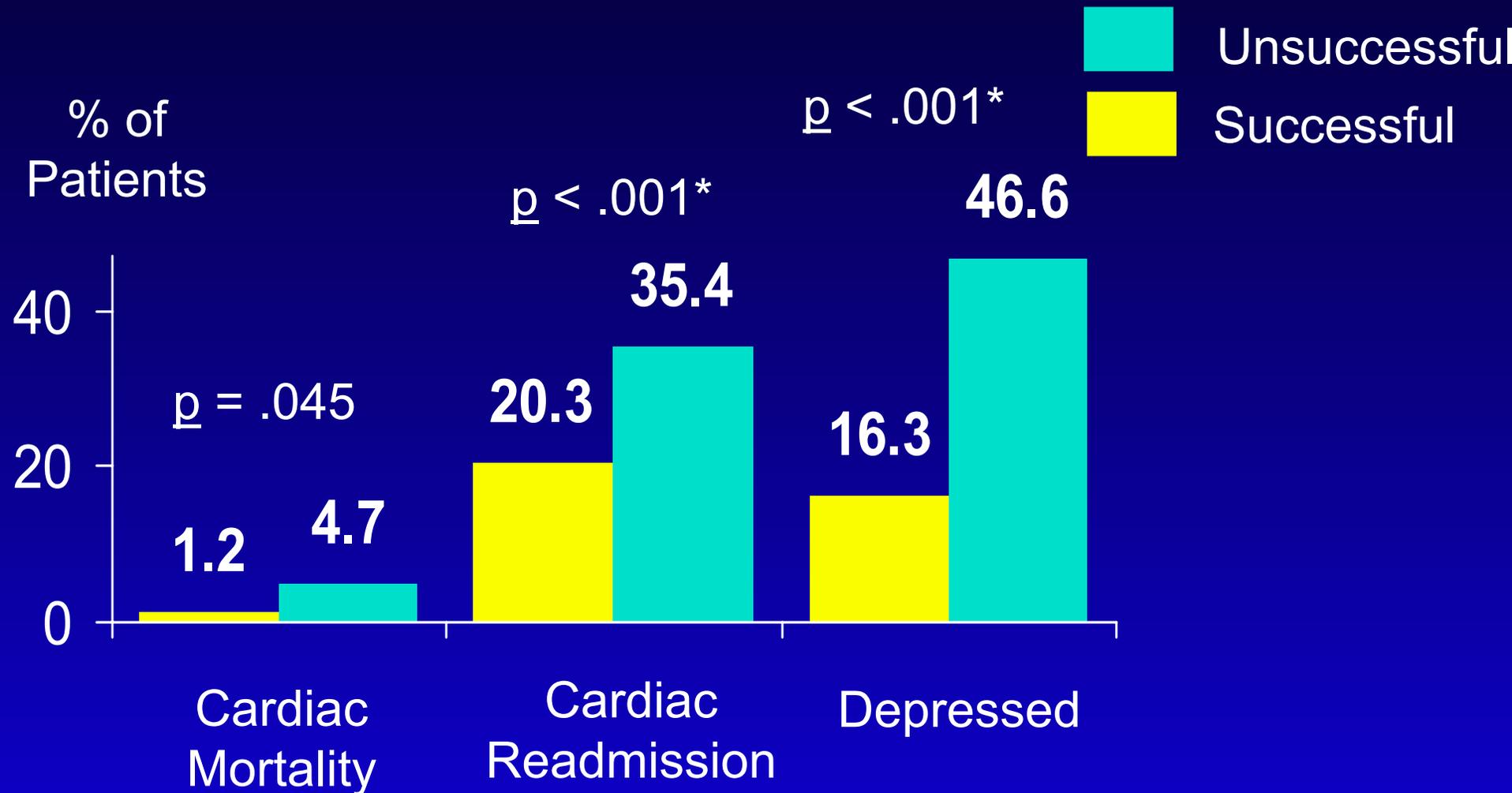
Impact of Improving Psychological Distress in MHART Patients

- Success = GHQ < 5 or > 50% reduction after two nursing visits
- Sample
 - 431 Treatment Group patients with 2 visits within a two-month interval after an initial elevated GHQ
 - 55.7% had successful GHQ outcomes

Baseline Characteristics Associated with Successful Reduction in GHQ after 2 Nurse Visits

	Successful	Unsuccessful	P-value
Men	68.2%	58.3%	0.029
Mean age	58.6	59.2	0.57
At least one Friend	81.2%	74.5%	0.090
First MI	78.8%	81.2%	0.53
Revascularized during Index	32.8%	21.4%	0.008
Not Depressed (BDI <10)	67.4%	52.7%	0.002
Not readmitted between 2 GHQ scores	89.2%	80.2%	0.009

Successful Short-term Changes in Distress Associated With Better 1-Year Cardiac and Psychological Outcomes Even after Adjustment for Covariates (*)



Conclusions

- Interventions of this type have the potential to reduce psychological distress in the short-term and improve long-term outcomes in post-MI patients
- Failure to alter psychological factors in enough patients could help explain the lack of impact of the M-HART study on post-MI prognosis
- Stepped care with evaluation of short-term outcomes, and reorientation of patients not showing early improvement might be an appropriate next step

Acceptable or not?

2) Ad hoc analysis of outcomes for treatment responders vs. non-responders

- Improper (potential bias) (-)
 - The responders might have gotten better anyway
- Provided the hypothesis of stepped care (+)

Subgroup Analyses in M-HART

- 3) Long-term ad hoc analysis to explore the possibility that the program may have interfered with normal coping in some patients, and paradoxically increasing distress

Low-anxious, High-anxious, and Repressive Coping Styles

(Weinberger et al, J Abnorm Psych, 1979)

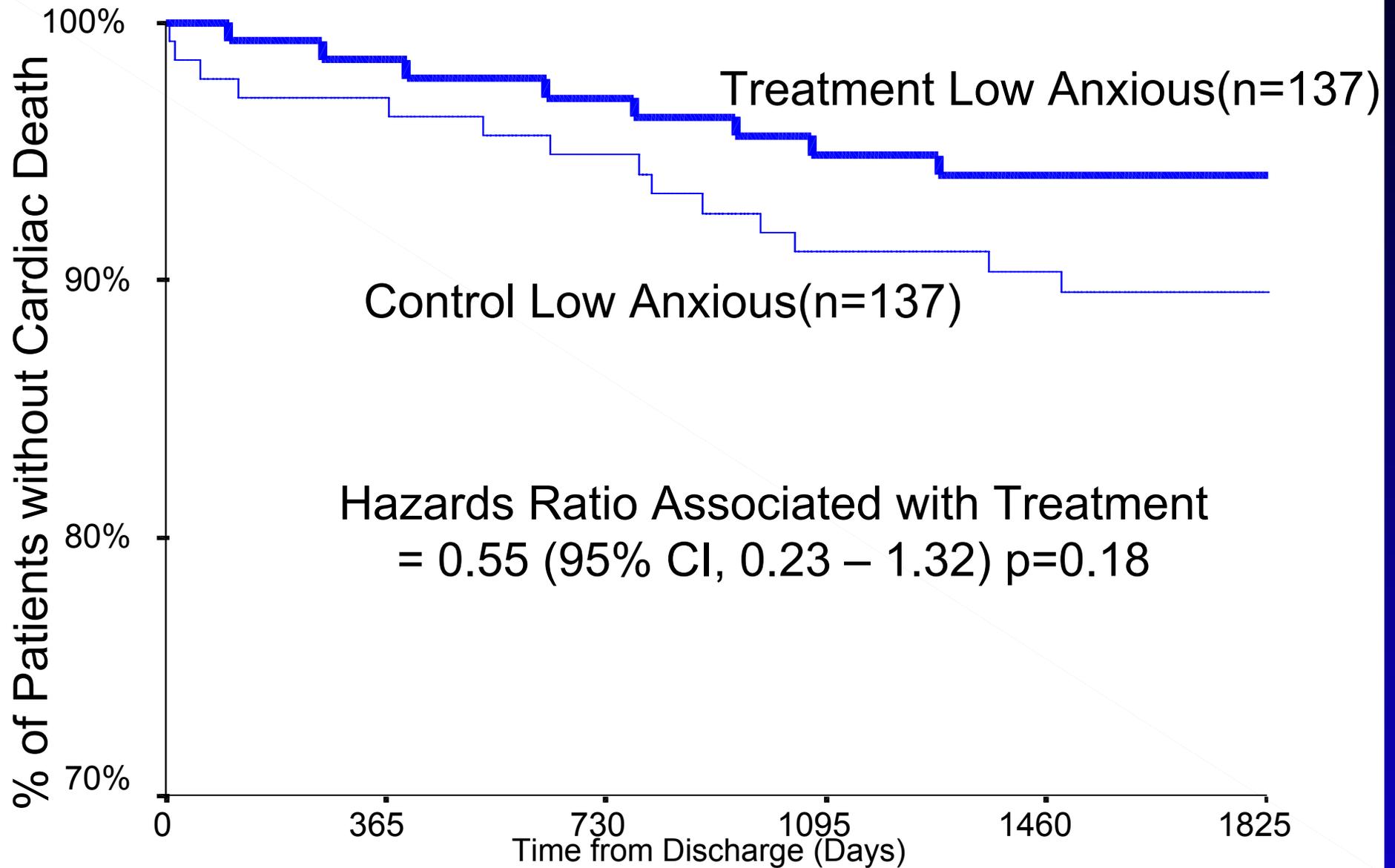
	Low Anxiety (S-STAI)	High Anxiety (≥ 35)
Low Marlowe-Crowne Social Desirability Score	Truly Low Anxious (n=274)	High Anxious (sensitizers; cope better with more information) (n=676)
High Marlowe-Crowne Score (> 9)	Repressors (generally cope better with low information) (n=408)	

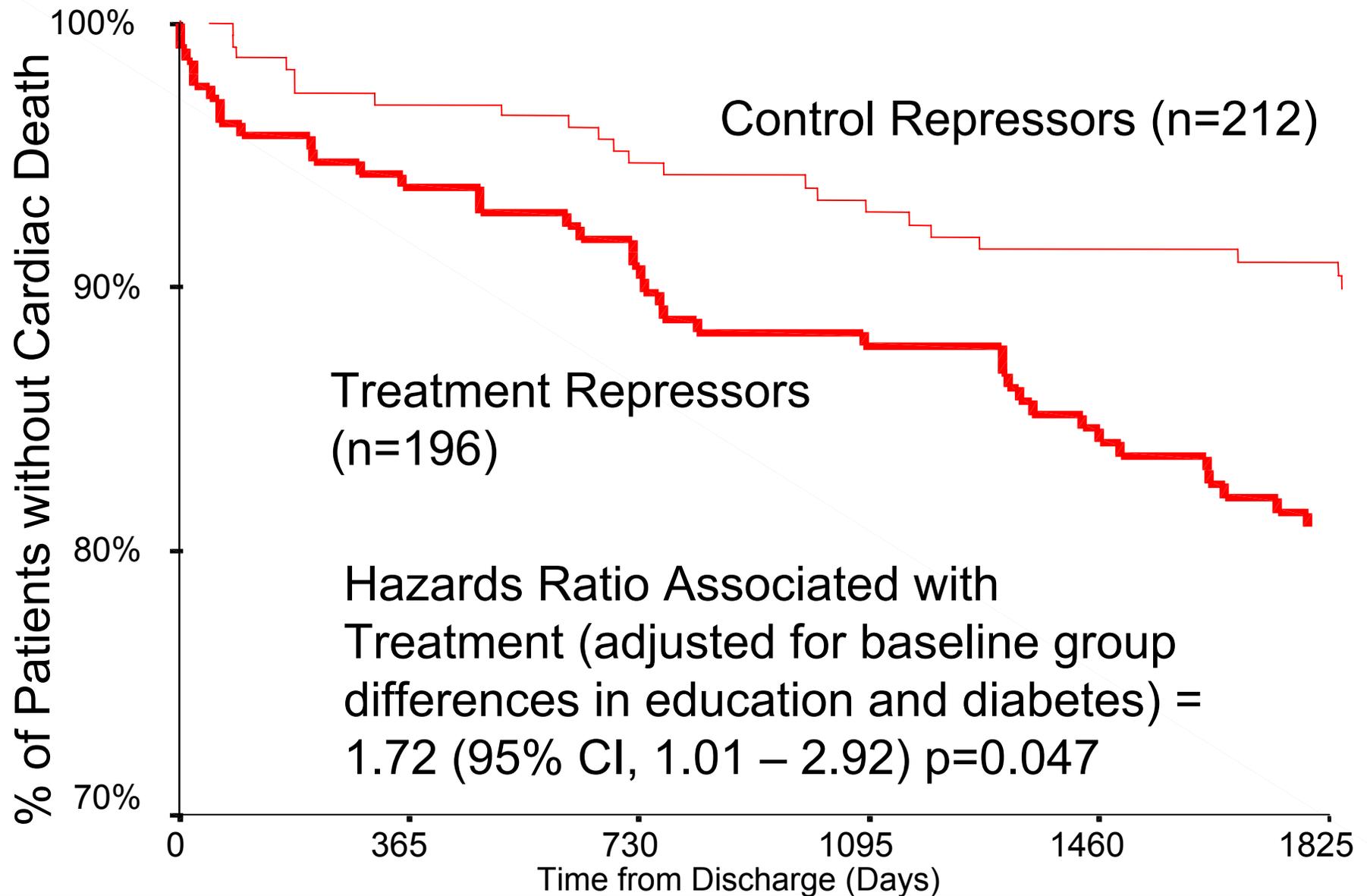
Results of Cox Proportional Hazards Regression Analysis for 5-year Cardiac Mortality

Sex	<0.001
Treatment Group	0.003
Coping Group	0.009
Treatment by Sex	0.002
Coping by Sex	0.013
Coping by Treatment	0.010
Coping by Treatment by Sex	0.015

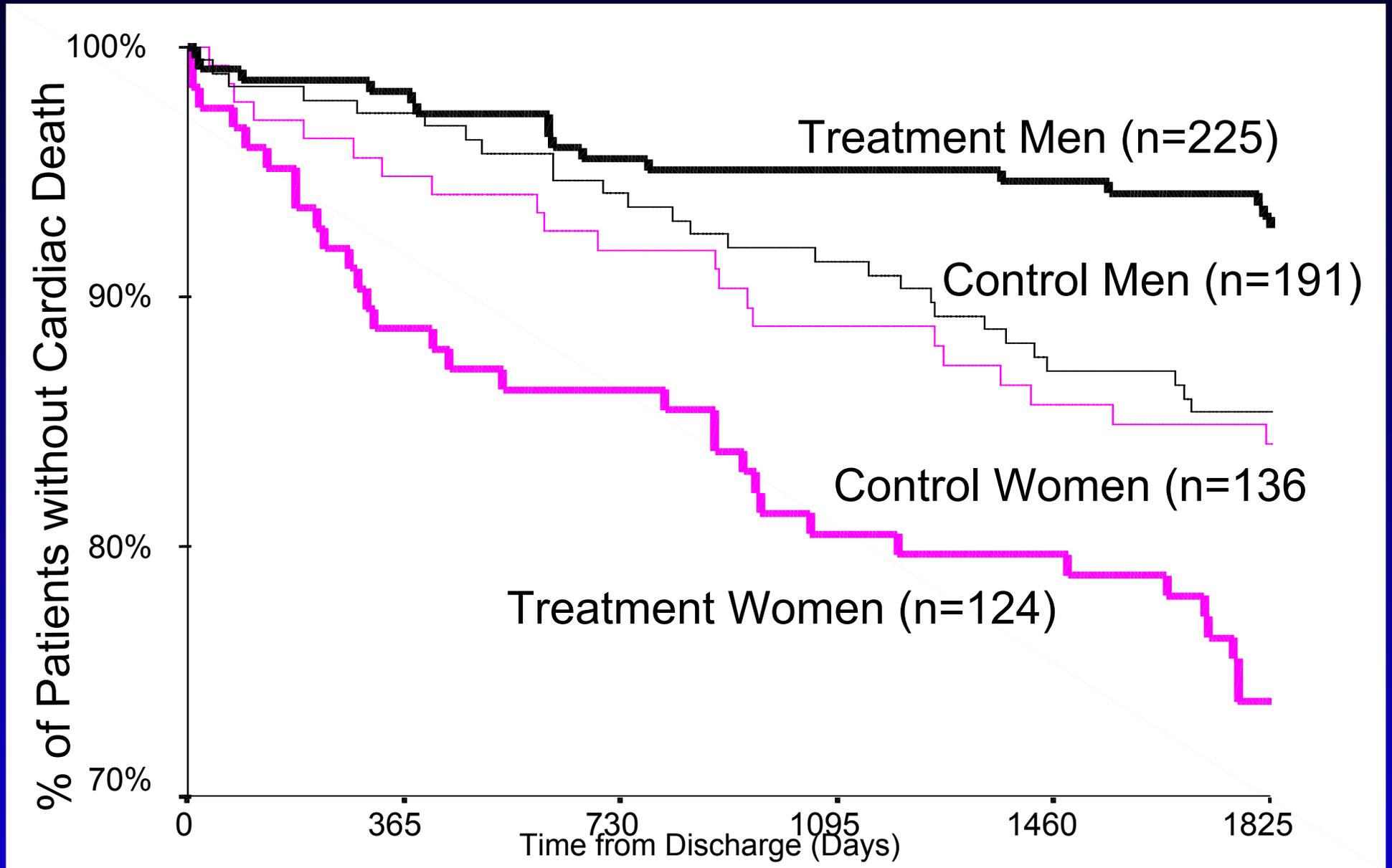
Results of 2- and 3-way Interactions for 5-year Cardiac Mortality

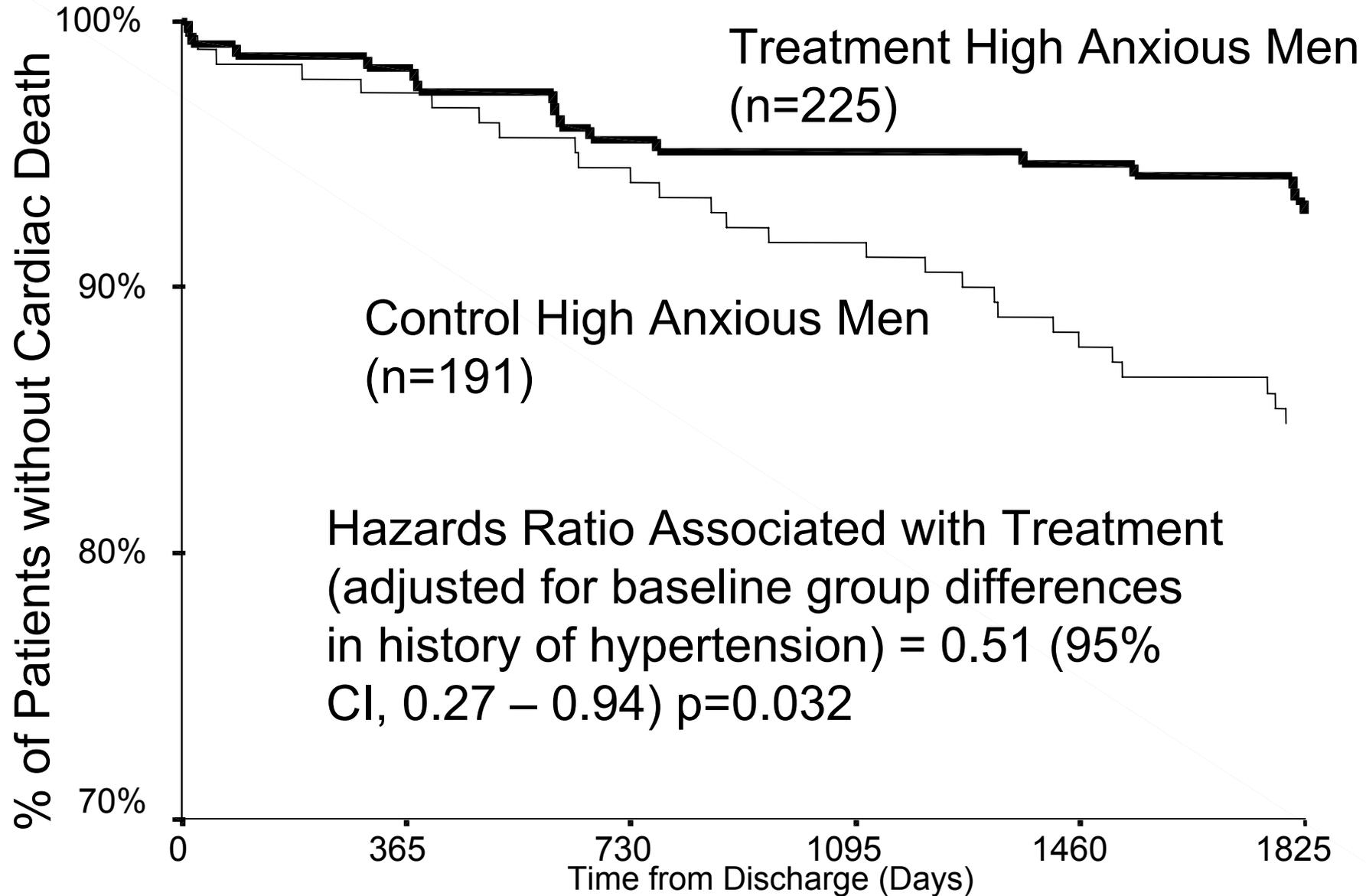
- Low anxious – No treatment impact
- Repressors – Negative treatment impact
- High Anxious
 - Men – Positive treatment impact
 - Women – No treatment impact





Treatment by Sex Interaction in High Anxious Patients





Search for Consistency in Evidence Using Mediator Analysis (Baron & Kenny, 1986)

- Variable linked to treatment group and to outcome
- Remains linked to outcome even after adjustment for treatment group
- Adjustment for mediator eliminates a large portion of the apparent link between treatment group and outcome

Potential Mediators for Treatment Outcomes in Repressors and High Anxious Men

- Differences in medical care and/or medications
 - Did repressors in the Treatment group rely on the nurses, and receive less attention from MDs than repressors in the Control group?
 - Was there more use of the health care system in the anxious men in the Treatment group than in the Controls?
- Differences in changes in negative emotions over year
 - Did high anxious men in the Treatment group show greater decreases in anxiety and depression symptoms?

Potential Medical Care Mediators for Worse Outcomes in Treated Repressors Surviving to 1-Year

Variable	T	C	P-value	HR for 5 yr CD	P-value
Mean MD Visits	13.6	12.8	0.44	1.6	0.083
% Any Psychiatrist Visits	1%	4%	0.10	0.1	0.47
% Any ER Visits without Admissions	37%	28%	0.044	1.9	0.040
% Any ER Visits Leading to Admissions	27%	25%	0.64	2.3	0.010
% Revascularized during Year	8%	15%	0.037	0.4	0.22
% Beta-blockers (1 year)	56%	58%	0.72	0.6	0.080
% Benzodiazepines	29%	16%	0.005	2.6	0.006
% Antidepressants	4%	2%	0.32	3.3	0.050

Potential Negative Emotion Mediators for Worse Outcomes in Treated Repressors

Variable (means adjusted for baseline values)	T	C	P-value	HR for 5 yr CD	P-value
Mean Depression Symptoms	5.0	4.3	0.24	1.05	0.061
Depressed (% BDI ge 10)	14.0	11.7	0.52	2.32	0.028
Somatic Symptoms	3.3	2.9	0.27	1.11	0.036
Cognitive Symptoms	1.7	1.4	0.41	1.05	0.27
Mean Anxiety Symptoms	30.1	28.8	0.16	1.01	0.57

Potential Mediators for for Worse Outcomes in Treated Repressors

- No evidence that changes in negative emotions over the program were involved, but by their nature, repressors are less likely to report on emotions
- Treatment repressors more likely to go to ER for problems that did not lead to admission, and more likely to be prescribed benzodiazepines by one year
- Control for each of these behaviors removed the apparent impact of treatment on prognosis
- Did the treatment increase distress in repressors?

Potential Medical Care Mediators for Better Outcomes in Treated High Anxious Men Surviving to 1-Year

Variable	T	C	P-value	HR for 5 yr CD	P-value
Mean MD Visits	16	15	0.31	1.1	0.75
% Any Psychiatrist Visits	10%	4%	0.016	0.4	0.42
% Any ER Visits without Admissions	41%	35%	0.24	1.0	0.99
% Any ER Visits Leading to Admissions	30%	29%	0.94	3.7	<0.001
% Revascularized during Year	15%	14%	0.79	0.2	0.080
% Beta-blockers (1 year)	53%	53%	0.93	0.3	0.027
% Benzodiazepines	25%	26%	0.80	1.6	0.25
% Antidepressants	4%	3%	0.59	1.0	0.99

Potential Negative Emotion Mediators for Better Outcomes in Treated High Anxious Men

Variable (means adjusted for baseline values)	T	C	P-value	HR for 5 yr CD	P-value
Mean Depression Symptoms	7.7	9.3	0.040	1.05	0.007
Depressed (% BDI ge 10)	27.6	37.5	0.042	2.22	0.042
Somatic Symptoms	3.9	4.8	0.015	1.17	<0.001
Cognitive Symptoms	4.0	4.5	0.28	1.04	0.18
Mean Anxiety Symptoms	36.8	39.3	0.036	1.00	0.99

Potential Mediators for Better Outcomes in Treated High Anxious Men

- No evidence that increases in medical care were involved
- Treated High Anxious men experienced a greater decline in depression symptoms than high anxious men in control group (particularly somatic symptoms of depression)
- Control for this change removed the apparent impact of treatment on prognosis
- Did the treatment decrease distress in high anxious men?

Long-term Subgroup Conclusions

- M-HART Program was not helpful for repressors of either sex
 - M-HART may have increased distress in repressors who tend to avoid information
- M-HART Program was not helpful for low anxious patients or high anxious women
- M-HART may have had a long-term positive outcome for high anxious men, the same group involved in the original IHD Life Stress Monitoring Program
 - M-HART may have reduced depression symptoms (somatic) in high anxious men

Acceptable or not?

3) Long-term ad hoc analysis to explore the possibility that the program may have interfered with normal coping in some patients, paradoxically increasing distress

- Proper subgroups (+)
- Hypothesis driven (+)
- Internally consistent (+)
- Many comparisons (-)
- $0 = +1 + (-1)$ (?)

Implications of M-HART and ENRICHD

- M-HART designed to lower distress, not to treat depression; ENRICHD targeted depression
- In M-HART, too many patients were treated for relatively low levels of distress (not helpful for repressors of either sex)
- M-HART may have had a long-term positive outcome for highly anxious men; the same group involved in the original IHD Life Stress Monitoring Program
- ENRICHD had long-term positive outcomes for depressed or socially isolated men

Implications of M-HART and ENRICHED

- M-HART was not appropriate for highly anxious women; ENRICHED was also not appropriate for depressed or socially isolated women
- What treatments should we provide for post-MI women?
 - Semi-qualitative analysis of nursing data suggests that women responded better to listening than to instructions and education (Cossette et al, Int J Nursing Studies, 2002)

What I learned about subgroup analyses (and you should remember)

- Pre-specified analyses are okay if limited in number
 - Try to assure adequate power
- Ad Hoc
 - For hypothesis generation only
 - More appropriate if there is an overall positive result
 - Systematic, data-driven analyses are problematic (false positives and negatives)
 - Preferably hypothesis-based, interpreted in the light of other studies

What I learned about subgroup analyses (and you should remember)

■ Proper

- Based on baseline characteristics

■ Improper

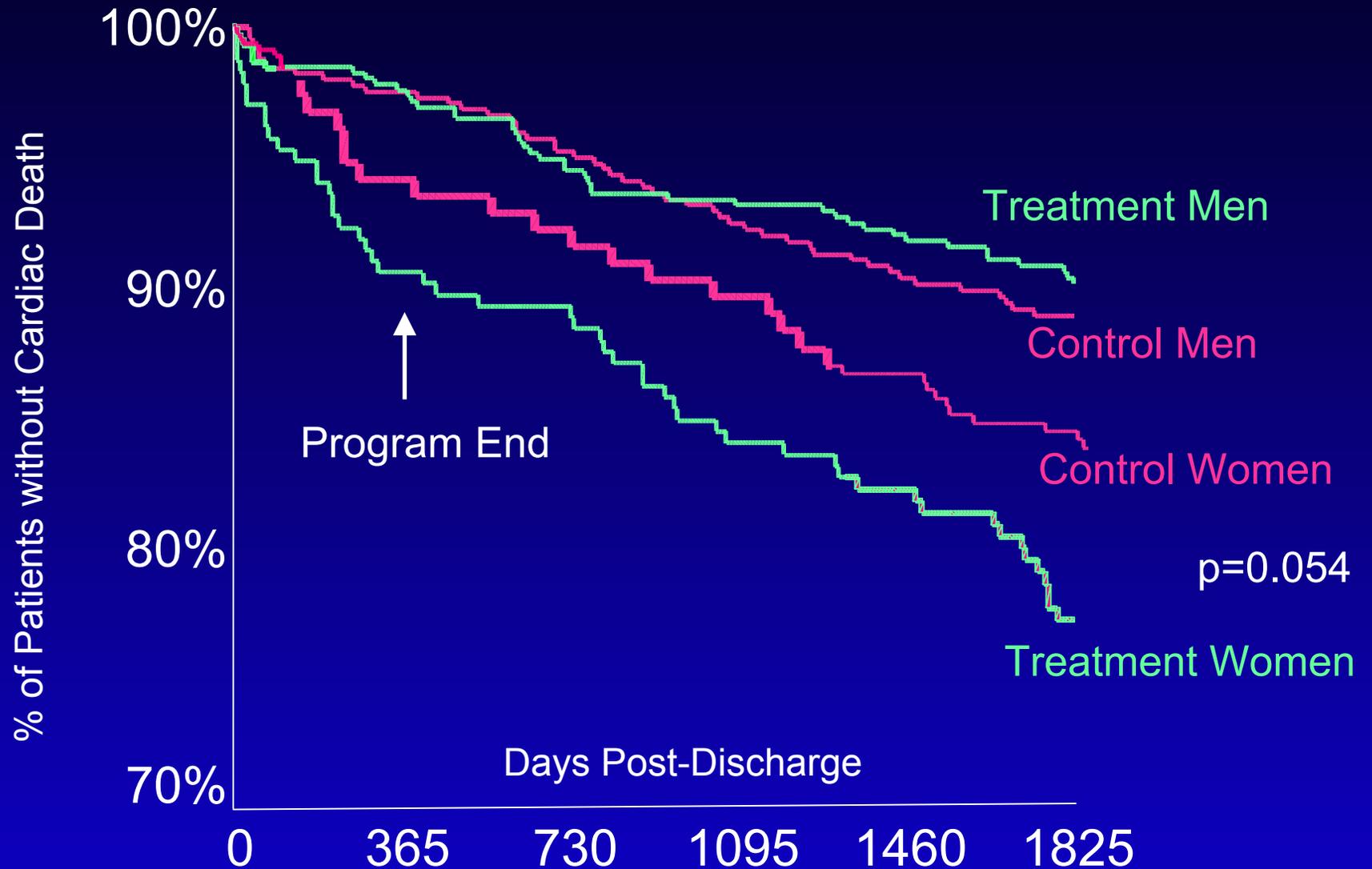
- Subgroups based on measurements later in the trial
 - Analysis of responders vs. non-responders



Subgroup
Analysis

~~Data
Dredging~~

5-Year Cardiac Mortality in M-HART



Results of Cox Proportional Hazards Regression Analysis for 5-year Cardiac Mortality

Coping by Treatment by Sex 0.015

Repressors (TxS) 0.95

Low Anxious (TxS) 0.11

High Anxious (TxS) 0.022

Coping by Treatment 0.010

Low Anxious 0.18

Repressors 0.013

High Anxious 0.71