Veterans Health Administration
Research & Development
Improving Veterans' Lives — www.research.va.gov

Complex Innovative Trial Designs in VA Cooperative Studies

Mei-Chiung Shih, Bruce Chow, Ilana Belitskaya-Levy
Palo Alto Cooperative Studies Program Coordinating Center

The Bay Area Biotech-Pharma Statistics Workshop
November 7-8, 2019

Outline

- **Brief Introduction of CSP**
- **Two Examples of Complex Innovative Trial Designs:**
  - **Adaptive multi-arm multi-stage design**
  - **Sequential randomized design**

VETERANS HEALTH ADMINISTRATION
Office of Research & Development

Chief Research & Development Officer:
Rachel Ramoni, DMD, ScD

<u>Research Services</u>
- Biomedical Lab
- Clinical Science
- Health Services
  - QUERI
- Rehabilitation

<u>Programs</u>
- **Cooperative Studies Program**
  (Director: Grant Huang)
- Million Veteran Program

<u>Policy & Operations</u>
- Finance
- Communications
- PRIDE
  - Central IRB
- Technology Transfer
- Non-Profits
- Bioinformatics

CSP Role

CSP designs and conducts multi-site cooperative research.

Ideas

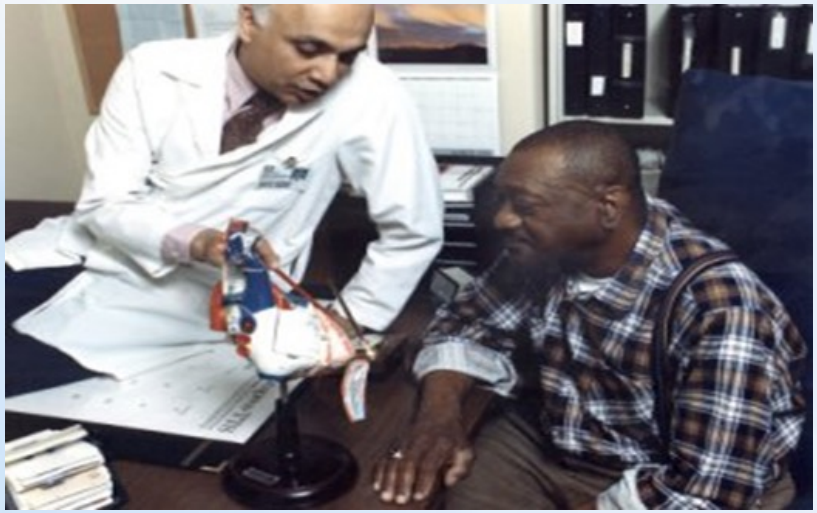
Resources

Tools

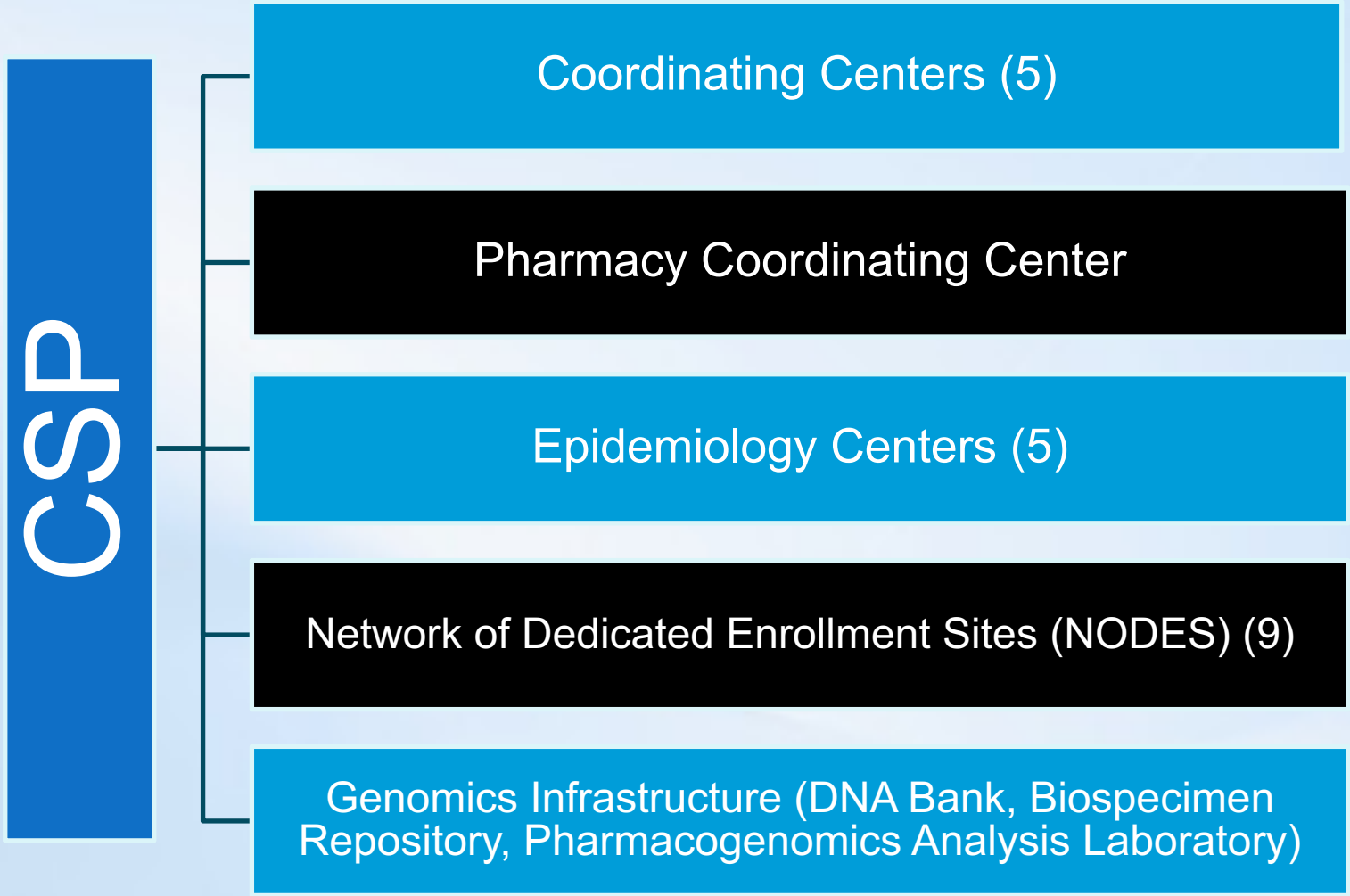
Policies & procedures

CSP Mission

To advance the health and care of Veterans through cooperative research studies that produce innovative and effective solutions to Veteran and national healthcare problems.

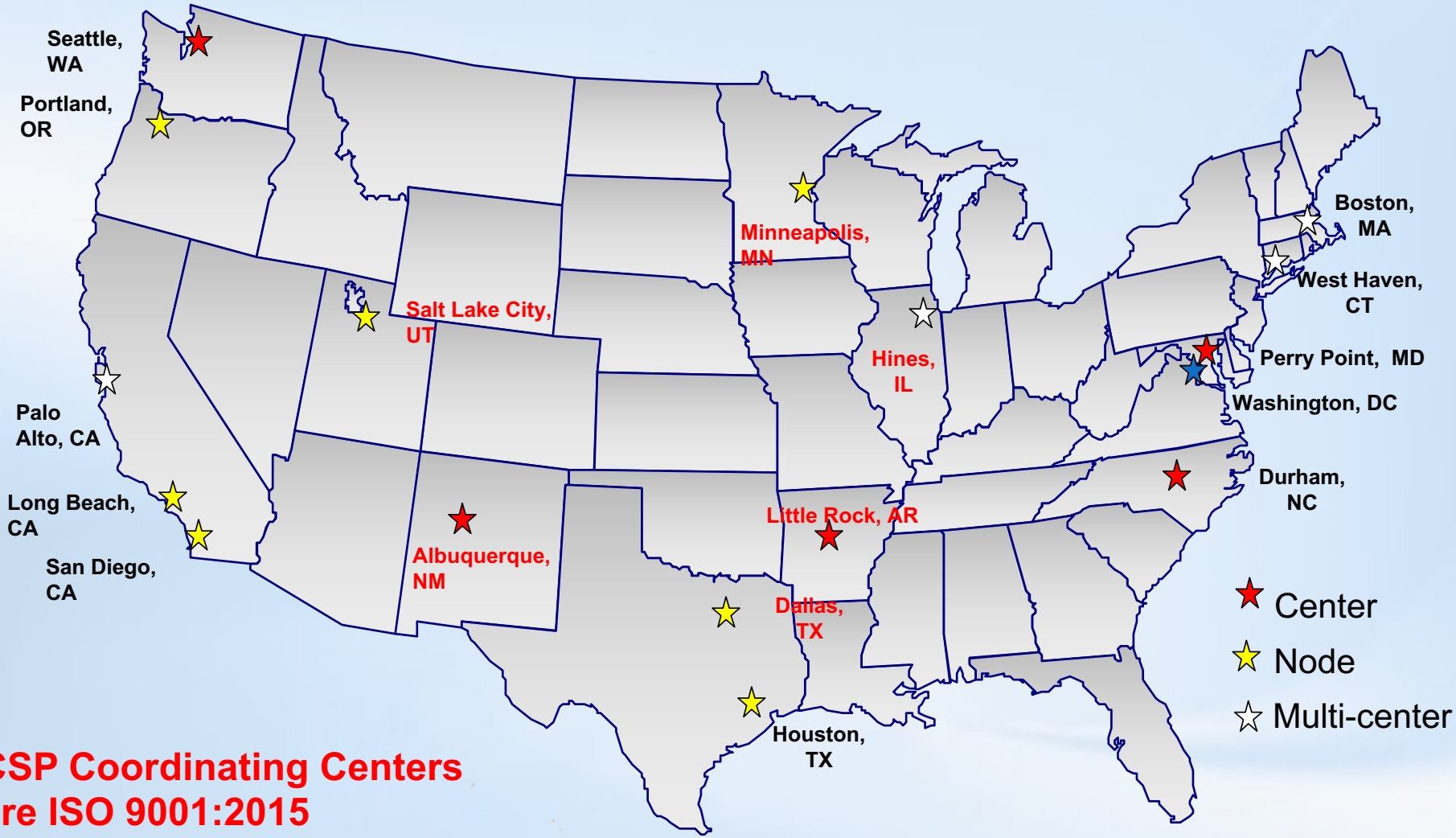


CSP Organization



CSP Center Locations

Veterans Health Administration
Research & Development
Improving Veterans' Lives ~ www.research.va.gov



CSP Coordinating Centers are ISO 9001:2015 registered

Learning Health Care in VA CSP

Veterans Health Administration
Research & Development

Comparative Effectiveness Research (CER)

- Direct comparison of existing health care interventions to determine the effectiveness, benefits and risks of different treatment options.
- Aim to support evidence-based choices of treatments for patients, providers, and health policy makers

CSP unique advantages

- Embed in the largest national health care system
- Community of ~3000 researchers
- >110 VAMCs have Federal Wide Assurances for research
- Central IRB
- Electronic Health Record since early 1980s
- Implementation of findings into VA and national healthcare

Example 1: CSP #2016

National Adaptive Trial for PTSD-related Insomnia (NAP)

Study Chair: John H. Krystal, MD

- ***Study objective*:** To evaluate the efficacy of ***trazodone hydrochloride***, ***eszopiclone*** and ***gabapentin*** (as compared to placebo) as an adjunctive therapy in the treatment of insomnia in veterans with military related PTSD
- *Primary outcome:* Change from baseline in Insomnia Severity Index at 12 wks
- Secondary outcomes: PTSD symptoms, depression, quality of life, etc
- Target sample size **1224**, to be recruited in 3 years from 34 sites
- Adaptive Multi-arm Multi-stage (MAMS) Design:
  - Plan to drop "non-promising" arms at the interim analysis
  - Re-allocate the remaining sample size of the stopped arms to other arms

CSP #2016 (NAP)

Veterans Health Administration Research Development

Considerations for interim analysis:

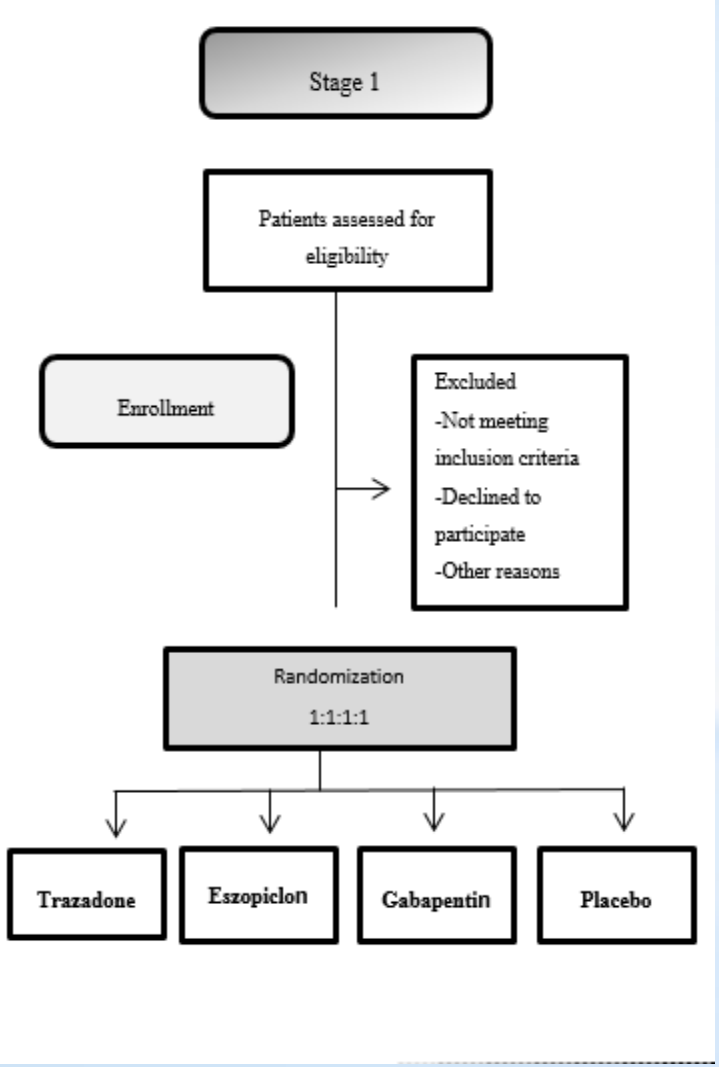
- Based on primary outcome
- Drop arms for futility?
  - Would there be sufficient evidence to change practice?
- Drop arms for efficacy?
  - Is it ethical to continue randomizing participants to placebo arm?
  - What about secondary outcomes (e.g., key secondary outcome CAPS-5 total score)?
- Binding vs non-binding
- Allocate remaining sample size of terminated arms to the remaining arms?
  - Pros: Increase the power for the remaining active arms; help maintaining interim analysis results
  - Cons: No saving in sample size; more complicated final analysis

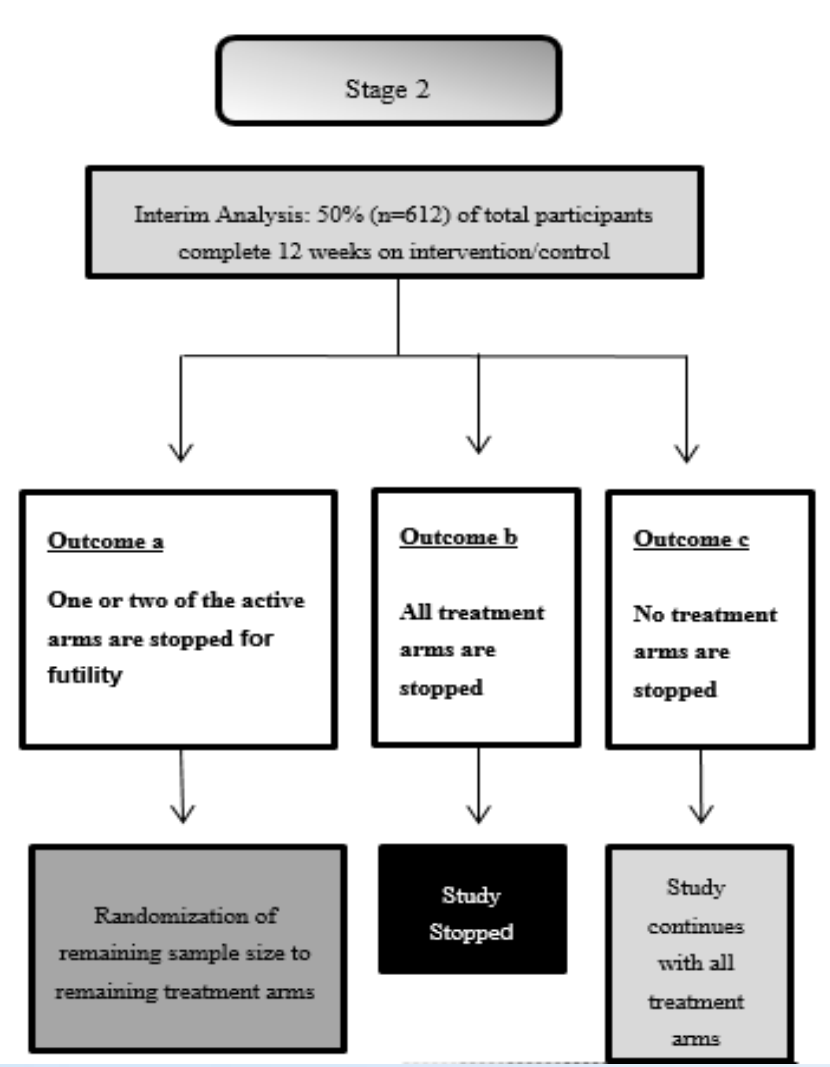
CSP #2016 (NAP)

Planned interim analysis:

- One single interim analysis when half of the target randomized participants have completed the 12-week follow-up (i.e., at 50% information time)
- An active treatment arm may be terminated for futility if there is sufficient evidence at the time of interim analysis that the effect size of the arm as compared to placebo is less than the clinically meaningful effect size of 0.35 in the primary outcome.
- More specifically, for each active treatment, we will perform a one-sided t-test of whether the effect size is less than 0.35 and terminate those active treatment arms which have p-value < 0.025.
- The study will be terminated at the interim if all active treatment arms are stopped early.
- Otherwise, allocate remaining sample size of terminated arms to the remaining arms.

CSP #2016 (NAP)


Stage 1
Patients assessed for eligibility
Enrollment
Excluded
-Not meeting inclusion criteria
-Declined to participate
-Other reasons
Randomization
1:1:1:1
Trazadone
Eszopiclon
Gabapentin
Placebo


Stage 2
Interim Analysis: 50% (n=612) of total participants complete 12 weeks on intervention/control
Outcome a
One or two of the active arms are stopped for futility
Outcome b
All treatment arms are stopped
Outcome c
No treatment arms are stopped
Randomization of remaining sample size to remaining treatment arms
Study Stopped
Study continues with all treatment arms

Sample size considerations

There are three sets of hypotheses for testing the efficacy of individual treatments:

$$H_{i0}: \delta_i \leq 0 \quad \text{versus} \quad H_{i1}: \delta_i > 0, \; i=1,2,3,$$

where δ_i is the mean difference in the primary outcome between treatment *i* and placebo, with positive values indicating bigger improvements in the active treatment arm.

- An effect size of 0.35 is clinically meaningful
- An effect size of 0.2 or smaller is not of clinical interest

Sample size considerations

The sample size of this study is selected to satisfy the following two requirements for power:

1) When all three active treatments have effect size 0.35, the study has at least 85% probability to establish efficacy of all three active treatments. *(Disjunctive power)*

2) When one active treatment (say, treatment 1) has an effect size 0.35 and the other two active treatments have an effect size 0.20, the study has at least 90% probability to recommend treatment 1 as an effective treatment. *(Power at the least favorable configuration)*

Operating Characteristics

Table 1.1 Probability of Stopping an active arm for futility at the interim analysis

Effect size	Probability that a study arm with the effect size is stopped for futility at the interim analysis using the proposed futility boundary
0.00	0.806
0.10	0.522
0.20	0.226
0.30	0.060
0.35	0.025
0.40	0.009
0.50	0.001

Operating Characteristics

Table 1.2 Operating characteristics of the futility boundary

Scenario	Effect size for the three active treatment arms	P(trt 1 is stopped at interim)	P(trt 2 is stopped at interim)	P(trt 3 is stopped at interim)	P(continue 3 active arms to stage 2)	P(continue 2 active arms to stage 2)	P(continue 1 active arm to stage 2)	P(study is stopped at interim)
1	(0,0,0)	0.807	0.814	0.805	0.046	0.104	0.228	0.622
2	(0.35,0.35,0.35)	0.024	0.025	0.027	0.937	0.053	0.010	0.001
3	(0.35,0.0,0.0)	0.025	0.801	0.803	0.084	0.228	0.664	0.025
4	(0.35,0.1,0.1)	0.025	0.512	0.520	0.317	0.332	0.330	0.022
5	(0.35,0.2,0.2)	0.025	0.224	0.230	0.648	0.238	0.100	0.013
6	(0.35,0.35,0.0)	0.023	0.025	0.812	0.188	0.768	0.039	0.004
7	(0.35,0.35,0.1)	0.024	0.025	0.521	0.475	0.484	0.036	0.005
8	(0.35,0.35,0.2)	0.023	0.023	0.223	0.763	0.209	0.025	0.003
9	(0.30,0.30,0.1)	0.060	0.058	0.527	0.460	0.450	0.076	0.014
10	(0.25,0.25,0.1)	0.120	0.123	0.525	0.438	0.396	0.127	0.039

Operating Characteristics

Table 1.3 Power of the study

Scenario	Effect size for the three active treatment arms	P(1)*	P(2)*	P(3)*	P(reject any H_{i0} where $\delta_i>0$)	P(reject all H_{i0} where $\delta_i>0$)	P(reject all H_{i0} where $\delta_i>0.2$ and none with $\delta_i\leq 0$)	FWER = P(reject any H_{i0} with $\delta_i\leq 0$)	Expected sample size
1	(0,0,0)	0.008	0.009	0.009	--	--	--	0.024	721
2	(0.35,0.35,0.35)	0.939	0.939	0.936	0.996	0.850	0.850	--	1039
3	(0.35,0.0,0.0)	0.971	0.012	0.012	0.971	0.971	0.948	0.023	1027
4	(0.35,0.1,0.1)	0.967	0.121	0.115	0.968	0.034	0.967	--	1029
5	(0.35,0.2,0.2)	0.950	0.465	0.464	0.964	0.281	0.950	--	1033
6	(0.35,0.35,0.0)	0.959	0.956	0.009	0.991	0.925	0.916	0.009	1037
7	(0.35,0.35,0.1)	0.954	0.954	0.104	0.991	0.103	0.918	--	1037
8	(0.35,0.35,0.2)	0.945	0.945	0.452	0.989	0.436	0.901	--	1037
9	(0.30,0.30,0.1)	0.875	0.876	0.112	0.965	0.107	0.786	--	1033
10	(0.25,0.25,0.1)	0.722	0.719	0.105	0.878	0.089	0.564	--	1018

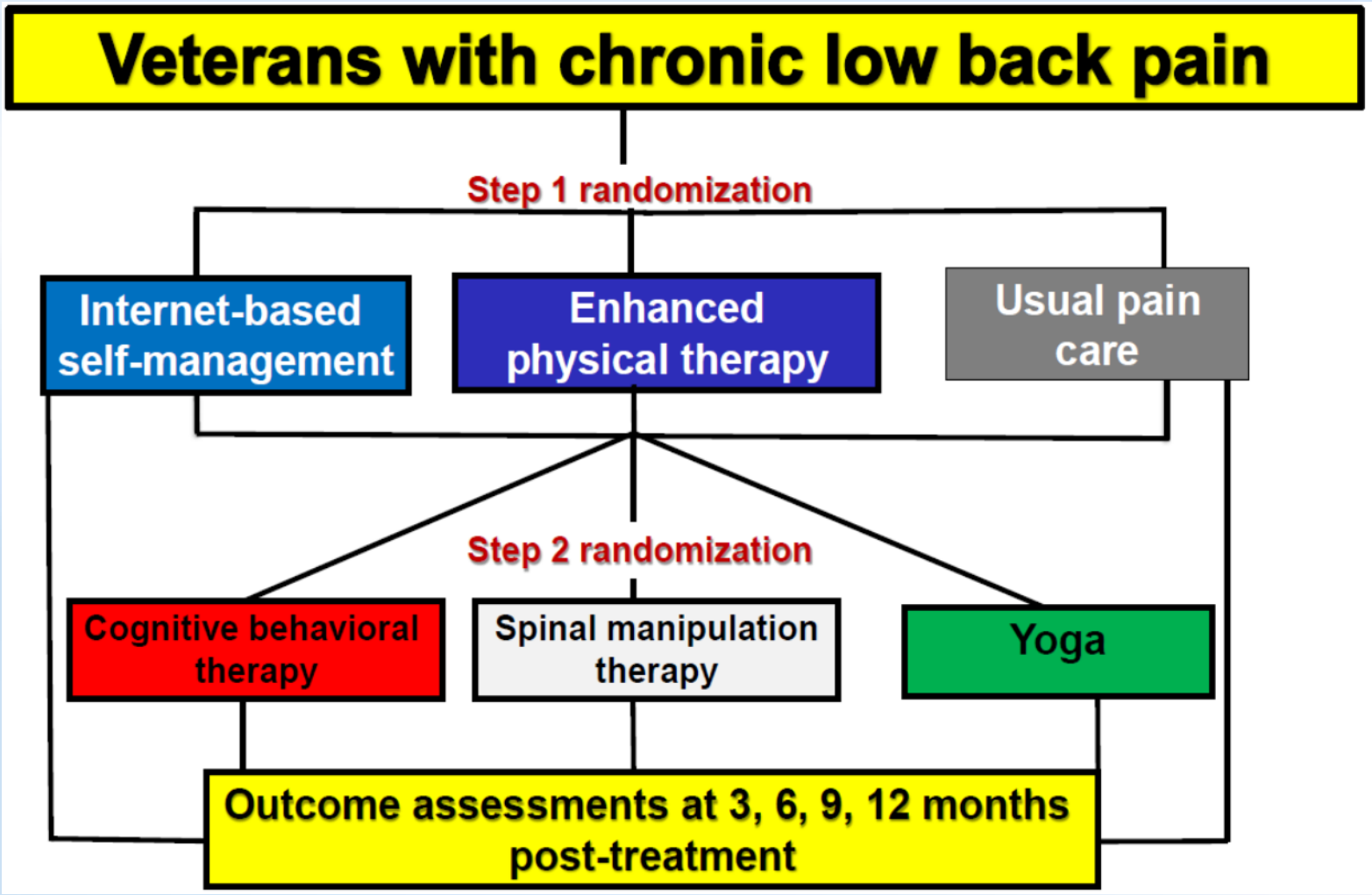
Example 2: CSP #2009

Sequential and Comparative Evaluation of Pain Treatment Effectiveness Response: The SCEPTER Trial

Study Chairs: David J. Clark, MD, PhD and Matthew J. Bair, MD, MS

- ***Study objective*:** To identify the optimal approach to chronic low back pain (cLBP) treatment employing commonly recommended non-surgical, non-pharmacological options.
- *Primary outcome:* Change from baseline in pain interference at 3 months
- Secondary outcomes/objectives:
  - Pain severity, depression, anxiety, sleep disturbances, quality of life, etc
  - Identify predictors of responses to treatments
  - Evaluate feasibility, barriers and facilitators to implementation
  - Cost and budget impact analysis
- Target sample size **2529**, to be recruited in 2.5 years from 20 sites

CSP #2009 (SCEPTER)



Primary Objectives

Primary objective 1:

To compare the effectiveness of an *internet-based pain self-management program* alone, an *enhanced physical therapy* intervention that combines pain self-management education with tailored exercise guided by a physical therapist, and *usual care* for the treatment of cLBP.

Primary objective 2:

To compare the effectiveness of *cognitive behavioral therapy*, *spinal manipulation therapy*, and *yoga* in Veterans without a clinically meaningful response to Step 1 treatment (Step 1 non-responders).

Sample size considerations

- A mean between-group difference of 1-point in the primary outcome (change of BPI interference score from baseline to 3 months) is considered a clinically meaningful difference
- Conservatively assuming a common SD of 2.5, a clinically meaningful difference of 1-point corresponds to a medium effect size of 0.4.
- The study sample size is determined by the number of Step 1 non-responders needed in Step 2 to have adequate power to detect all pairwise comparisons among the Step 2 treatments that have a mean group difference of 1-point or greater in the primary outcome.

Sample size considerations

Table 2.1 Number of Step 1 non-responders needed at Step 2 to achieve desired power to detect all pairwise differences between Step 2 treatments

SD	Effect size corresponding to 1-point difference	Mean reduction for the 3 arms	All-pairs power		
			80% power	85% power	90% power
2.5	0.40	(1, 2, 2)	153 per arm	172 per arm	194 per arm
		(1, 2, 3)	166 per arm	180 per arm	**202 per arm**
3.0	0.33	(1, 2, 2)	222 per arm	245 per arm	277 per arm
		(1, 2, 3)	237 per arm	257 per arm	288 per arm

Sample size considerations

The final sample size was derived after adjusting for

- Correlation due to therapist
- Equipoise randomization at Step 2
- Anticipated proportion of Step 1 non-responders who are willing to proceed to Step 2
- Drop out rate

Sensitivity Analysis for Power

Table 2.2 Sensitivity analysis for the power of primary objective 2

SD	Percent of Step 1 participants with primary outcome who are non-responders and willing to proceed to Step 2					
	60% (inflation factor 1.72)		50% (Inflation factor 1.57)		40% (inflation factor 1.45)	
	Mean reduction in BPI in the 3 treatment arms					
	(1, 2, 2)	(1, 2, 3)	(1, 2, 2)	(1, 2, 3)	(1, 2, 2)	(1, 2, 3)
2.0	99%	99%	99%	98%	97%	97%
2.5	92%	90%	88%	87%	82%	78%
3.0	74%	69%	70%	61%	60%	49%

Power for Primary Objective 1

Table 2.3 Power to detect pairwise differences among the Step 1 treatments

	Enhanced Physical Therapy versus usual care or versus internet-based self-management program				Usual care versus internet-based self-management program
Effect size	ICC $\rho = 0.02$	ICC $\rho = 0.03$	ICC $\rho = 0.04$	ICC $\rho = 0.05$	ICC $\rho = 0.00$
0.15	62%	57%	54%	50%	72%
0.20	88%	85%	82%	78%	94%
0.25	98%	97%	96%	94%	99%
0.30	>99%	>99%	99%	99%	>99%

CSP #2009 (SCEPTER)

Considerations for interim analysis:

- Drop Step 1 treatment arms?
- Drop Step 2 treatment arms?
- Drop treatment strategies?

It was decided that the study does not plan to stop treatment arms early for efficacy for futility.

CSP #2009 (SCEPTER)

Rationale for not stopping treatment arms early for efficacy or for futility

- **Step 1:**
  1) There are no ethical concerns;
  2) It is important to examine durability of treatment effect in Step 1 responders and comparative effectiveness of subsequent Step 2 treatments in Step 1 non-responders even when there is no difference between Step 1 treatments at 3 months;
  3) It allows us to examine potential interactions between Step 1 and Step 2 treatments.

CSP #2009 (SCEPTER)

Rationale for not stopping treatment arms early for efficacy for futility

- **Step 2:**
  1) There are no ethical concerns
  2) Even when there are differences between these treatments, the differences based on interim analysis results are not likely to alter VA policy to make all evidence-based treatments available to Veterans with cLBP.
  3) Need to collect sufficient data on the secondary outcomes to support findings in the primary outcome.
  4) It allows us to examine the impact of patient preferences, treatment expectations, and patient characteristics on treatment effectiveness.
  5) It allows us to explore adaptive treatment strategies of cLBP.

Concluding Remarks

- Complex innovative trial designs can be useful in pragmatic comparative effectiveness setting
- Other examples include outcome adaptive randomization and contextual multi-arm bandit design
- Thanks to CSP #2009 and #2016 study teams and study planning committees