

STATISTICAL ANALYSIS PLAN

Official Title

Renewal of the Multi-center Randomized Controlled Trial of Refeeding in Anorexia Nervosa

Study

Individualized Study of Refeeding to Optimize iNpatient Gains (iStRONG)

ClinicalTrials.gov Identifier

NCT04966858

Document Date

May 6, 2026 (Version 1.0)

iStRONG Study Protocol and Statistical Analysis Plan (SAP)

MULTI-CENTER RANDOMIZED CONTROLLED TRIAL OF REFEEDING IN ATYPICAL ANOREXIA NERVOSA:

The Individualized Study of Refeeding to Optimize iNpatient Gains (iStRONG)

Principal Investigator	Andrea Garber, PhD, RD Professor of Pediatrics Division of Adolescent & Young Adult Medicine UCSF Benioff Children's Hospital University of California, San Francisco 550 16 th St San Francisco, CA 94143 ph: 415-514-2180
Former Multiple Principal Investigator	Neville Golden, MD Stanford Principal Investigator, 2020–2024 Division of Adolescent Medicine The Marron and Mary Elizabeth Kendrick Professor of Pediatrics Emeritus-Active Stanford University School of Medicine 770 Welch Road, Suite 100 Palo Alto, CA 94304
Stanford Site Principal Investigator	Cynthia Kapphahn, MD Stanford Site Principal Investigator, 2025 Division of Adolescent Medicine Stanford University School of Medicine
Protocol identification number	2R01HD082166-06A
UCSF IRB	20-32551
ClinicalTrials.gov identifier	NCT04966858
Author	Andrea Garber
Version (Date)	[1.0] (May 6)
Data Coordination Center	Jing Cheng, PhD ; Senior Statistician Sasha Gorrell, PhD ; Lead Analyst

Abbreviations

Abbreviation	Definition
iStRONG	individualized Study of Refeeding to Optimize iNpatient Gains (current trial)
StRONG	Study of Refeeding to Optimize iNpatient Gains (parent trial)
AAN	Atypical anorexia nervosa
AN	Anorexia nervosa
DCC	Data Coordination Center
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, fifth edition
ICR	Individualized calorie refeeding
HCR	Higher calorie refeeding
LCR	Lower calorie refeeding
kcal	Kilocalories
mBMI	Median Body Mass Index (for age and sex)
%mBMI	% of Median Body Mass Index
SOC	Standard of Care
HR	Heart rate
SBP	Systolic blood pressure
BP	Blood pressure
IRB	Institutional Review Board
HIPAA	Health Insurance Portability and Accountability Act
EMR	Electronic medical record
T3	Triiodothyronine
E2	Estradiol
LH	Luteinizing hormone
FSH	Follicle-stimulating hormone
REE	Resting energy expenditure
RMR	Resting metabolic rate
VO2	Volume of oxygen consumed
VC02	Volume of carbon dioxide produced
PHQ-A	Patient Health Questionnaire, Adolescent Version (replaced with PHQ-8)
PHQ-8	Patient Health Questionnaire, 8-item
WS	Weight suppression
EDE	Eating Disorder Examination
EDE-Q	Eating Disorder Examination-Questionnaire

WBIS-M	Modified Weight Bias Internalization Scale
SATAQ-4R	Sociocultural Attitudes Towards Appearance Questionnaire-4 Revised
BI-AAQ	Body Image-Acceptance and Action Questionnaire
GAD-7	Generalized Anxiety Disorder 7-item scale
EDE-15-Y	Eating Disorder 15 for Youth
ED-15	Eating Disorder 15 (adult version)
DTS	Distress Tolerance Scale
AARP	Treatment Acceptability and Appropriateness measure
PAQ	Physical Activity Questionnaire
CRAFFT	Car, Relax, Alone, Forget, Friends, Trouble (substance use screen)
SUDS	Subjective Units of Distress Scale
PVED	Parents Versus Eating Disorders
SSS	Satisfaction with Services
PVA	Parents Versus Anorexia Scale
PEDE-Q	Parent Eating Disorder Examination-Questionnaire
PED-15	Parent Eating Disorder 15 (caregiver report)
AE	Adverse event
DSMB	Data Safety Monitoring Board
OnCore	Online Collaborative Research Environment (clinical trial management system)

Table of Contents

MULTI-CENTER RANDOMIZED CONTROLLED TRIAL OF REFEEDING IN ATYPICAL ANOREXIA NERVOSA:	1
The Individualized Study of Refeeding to Optimize iNpatient Gains (iStRONG)	1
Abbreviations	2
Table of Contents	4
STUDY PROTOCOL (Version 1.0)	5
1. Introduction	5
2. Study Design	5
2.2. Randomization	6
3. Study Procedures	6
3.2. Safety	9
STATISTICAL ANALYSIS PLAN	11
1. Aims and Objectives	11
2. Statistical Methods	12
2.1. Pool of Participants	12
2.2. Outcomes	12
2.3. Power and Sample-Size Considerations	15
2.4. Data Management	16
2.5. Retention and Attrition	17
APPENDIX	18
1. Amendments to the Protocol	18
2. Amendments to the Statistical Analysis Plan	19

STUDY PROTOCOL (Version 1.0)

1. Introduction

The iStRONG study focuses on hospitalized adolescents and young adults diagnosed with atypical anorexia nervosa (AAN). The study aims to compare the efficacy, safety, and clinical remission between higher-calorie refeeding (HCR) and individualized caloric refeeding (ICR). The prior StRONG study demonstrated the safety and efficacy of HCR as compared to low calorie refeeding (LCR). HCR is now considered the new standard of care. However, subsequent analyses revealed that patients with AAN gained less weight, required longer to restore medical stability and stayed longer in hospital, even when refeed via HCR. Exploratory moderator analyses suggested that these poor outcomes were due to underfeeding. We subsequently developed an individualized caloric refeeding (ICR) approach that “doses” caloric level based on body weight: 50 calories per kg of body weight was selected because it was associated with optimized outcomes in patients with AN. The iStRONG trial was designed to compare these two treatments in a parallel, randomized fashion.

2. Study Design

The purpose of this multi-center RCT is to compare ICR vs. HCR in hospitalized adolescents and young adults with AAN. The trial is powered to detect a meaningful difference in time to restore medical stability (as defined below). Briefly, participants will be recruited upon hospital admission at UCSF and Stanford and randomly assigned to ICR or HCR, for a total planned sample of N=74. Treatments are not blinded, since both the patients and clinicians who work with this population are highly skilled at estimating kcal and would be able to determine group assignment by simply viewing meal trays.

2.1. Study Population

- a. *Inclusion/Exclusion:* Inclusion criteria are age 12-24 yrs, DSM-5 criteria for AAN, and hospitalized with medical instability: 1.) nighttime heart rate (HR) <45 bpm and/or daytime HR <50 bpm, 2.) systolic blood pressure (SBP) <90 mmHg, 3.) Temperature <35.6° C, 4.) orthostatic Δ HR >35 bpm, and 5.) orthostatic Δ SBP >20 mmHg. Patients with bulimia nervosa; current pregnancy, chronic disease (e.g., immune, renal disease), acute suicidality or psychosis, or hospital admissions for refeeding in the prior 6 mo. will be excluded.
- b. *Participant recruitment and consent:* Participants and their parents will sign assent (for those < 18 yr) and consent, respectively within 24 hrs of hospitalization. This may occur in the clinic when they are deemed medically unstable and waiting transfer to the hospital or in the hospital if admitted directly. Consent will include permission to review all medical records, to review hospital billing data, and to contact for future research projects. If a participant turns 18 yr while enrolled, s/he will be reconsented. Participants 18 years of age and older are able to consent themselves; thus, we will request verbal consent from their parents to complete parent surveys.

2.2. Randomization

Participants will be stratified by site and randomly assigned within site 1:1 to ICR or HCR within 24 hrs of hospitalization. The Data Coordination Center (DCC) will provide a secure, unpredictable allocation sequence (e.g., A B B A ...), which will be programmed into a secure electronic study tracking system for the assignment of each accrued participant. The sequences will be generated using a block size of two to four to maximize the balance between arms throughout accrual while ensuring the sequences remain unpredictable. As patients consent to study participation, clinical research staff will assign the following available study ID number in sequence, identify the allocated intervention arm, and inform participants and their families of the assignment. In turn, the clinical research staff will provide the linked study ID number and Medical Record Number to the DCC, storing this Personal Health Information in a HIPAA-compliant manner along with the intervention assignment.

3. Study Procedures

3.1 Treatment, Follow-Up, and Data Collection

- A. **Study groups:** Group A will receive ICR, beginning with 50 kcal/kg; group B will receive HCR, beginning at 2000 kcal/d; both groups will advance 200 kcal/d.
- B. **Refeeding intervention:** Our refeeding protocols are published and have been reproduced. Briefly, meal-based refeeding includes 3 meals + 2-3 snacks/day, on bedside trays, in the presence of meal monitors who observe intake of all nutrition and remain in the room for 20-30 min afterwards. Nasogastric tube feeding is rare; patients may drink a liquid supplement (formula) providing 1.5 kcal per mL (360 kcal per 240 mL can) to replace uneaten/refused kcals. Pre-filled diet orders for ICR or HCR are embedded in the EMR, signed by the attending physician, and advanced per protocol. Meals are prepared by hospital foodservice per U.S. Dietary Guidelines (~30-40% fat, 15-25% protein and 35-55% carbohydrate). Menus are continually analyzed via the hospital foodservice database (Software v.19.4, Computrition, Inc., Westhill, CA) to ensure conformance to this distribution as menu items are added or changed over time.
- C. **Blood draws (5 minutes):** We are proposing six blood draws: twice in the hospital and four times in follow-up. Daily, fasting blood draws in the hospital between 5-7 AM are Standard of Care (SOC) for the first 7 days, every other day thereafter, and more frequently if needed. These electrolyte monitoring blood samples provide the safety data for AIM 2. Additional blood samples for biomarkers will be collected in the hospital by adding two additional tubes to the SOC blood draw (12 mL) performed within 48 hrs of admission and discharge, thus avoiding additional venipuncture. To be consistent with the inpatient protocol, follow-up blood draws will be performed in the morning while fasting. All samples will be drawn and processed via a standardized protocol in the clinical laboratories at UCSF and Stanford. Both are equipped with research cores staffed by phlebotomists trained in research specimens. Samples will be centrifuged, separated as serum and plasma

into eight 0.5 mL aliquots, stored at -80 °C and sent for batched analysis. Testosterone will be assessed with Enzyme-linked immunosorbent assay (ELISA); cortisol, E2, LH, FSH, and T3 are assessed by chemiluminescent immunoassay; leptin is assessed by radioimmunoassay.

- D. **Indirect calorimetry (50 min including rest):** Resting energy expenditure (REE) will be measured in the morning, while fasting and awake, following the blood draw and while lying comfortably for 30 min. We will use an ambulatory indirect calorimeter that is validated against stationary equipment. This equipment calibrates on room air, avoiding the need for air tanks that prevent transport. It also features a mouthpiece and nose clip instead of an enclosed breathing canopy, which improves comfort. REE is measured from expired air using a 5 L mixing chamber technique, which samples every 15 seconds. A 6-ft breathing tube connects the non-breathing valve to the mixing chamber inlet. VO₂ is calculated using modified Haldane equations, whereas CO₂ is directly measured by the CO₂ analyzer within the device. Output includes RMR, VO₂ and VCO₂ with a calculated Respiratory Quotient to examine fuel use (percent of carbohydrate vs. fat). Data are provided every 15 seconds for the duration of the test with a graph plotting VO₂, CO₂ and RMR over the duration of the test.
- E. **Records and questionnaires (100 min at baseline, 40 min at follow-ups):** Collected in hospital and follow-up as specified. Questionnaires are administered securely online via Qualtrics and completed prior to or during study visits on laptop, phone, or tablet.
- Demographics, eating disorder history and physical (15 min, baseline only):** baseline information from the SOC physician examination and proctored questionnaire includes weight suppression (WS), pubertal stage, menstrual history, medications (including hormonal contraceptives), family history of eating disorder, self-reported race/ethnicity, maternal education and zip code (socioeconomic status), date of birth.
 - Growth records:** Growth records (height, weight and BMI curves) are obtained from Primary Care Providers as part of the referral packet to our eating disorder programs and used to evaluate weight history, WS, and other growth patterns that may predict response to refeeding and recovery targets.
 - Healthcare utilization form (5 min):** medications, mental health care, other care outside of our medical centers (e.g., residential care), and readmissions collected as potential prognostic covariates of long-term outcomes in this open follow-up study, and to assess cost-efficacy of the proposed treatment.
 - Patient Health Questionnaire, 8-item (PHQ-8, 2 min):** an 8-item depression severity measure (items scored 0-3, range 0-24); does not include a suicidality screening item. Administered at admission, 3 months, 6 months, 9 months, and 12 months.

- e. **EDE-Q (5 min)**: A self-report measure of eating disorder psychopathology over 28 days, validated against clinical interview and used as a recovery outcome. Administered at all timepoints (admission, discharge, 3 months, 6 months, 9 months, and 12 months). Parent-EDE-Q (PEDE-Q, 7 min, parent/caregiver only) is a secondary measure often yielding greater severity than endorsed by the adolescent; administered at the same timepoints.
- f. **Modified Weight Bias Internalization Scale (WBIS-M, 2 min)**: measure of the degree to which individuals internalize negative stereotypes about their weight status; relevant in AAN patients who may have experienced stigma at their pre-morbid higher weight. Administered at admission, 3 months, 6 months, 9 months, and 12 months.
- g. **Sociocultural Attitudes Towards Appearance Questionnaire-4 Revised (SATAQ-4R, 5 min)**: measure of societal influences on body image and eating disturbances and internalization of appearance ideals; relevant in AAN given higher weight individuals' nonconformity to the "thin ideal" and societal pressure. Administered at admission, 3 months, 6 months, 9 months, and 12 months.
- h. **The Body Image-Acceptance and Action Questionnaire (BI-AAQ, 2 min)**: a measure of experiential avoidance of and flexibility with body image; relevant in AAN because body acceptance may be more challenging for higher weight individuals. Administered at admission, 3 months, 6 months, 9 months, and 12 months.
- i. **Generalized Anxiety Disorder-7 (GAD-7, 2 min)**: self-report questionnaire used to screen for and assess the severity of anxiety symptoms. Administered at admission, 3 months, 6 months, 9 months, and 12 months.
- j. **Eating Disorder 15 for Youth (EDE-15-Y, 2 min)**: a brief, 15-item questionnaire to assess eating disorder symptoms in young people ages 8 and older, adapted from the adult ED-15 questionnaire. Administered at all timepoints (admission, discharge, 3 months, 6 months, 9 months, and 12 months). Parent ED-15 (PED-15, parent/caregiver report) is administered at the same timepoints.
- k. **Physical Activity Questionnaire (PAQ, 5 min)**: adapted questionnaire to assess type and time spent in physical activities and activities of daily living in the week prior. Administered at admission, 3 months, 6 months, 9 months, and 12 months.
- l. **CRAFFT (1 min)**: a series of 6 questions to assess alcohol, drug, and other substance use in adolescents and young adults. Administered at admission only.

- m. **Diet recall and calorie counts (20 min):** 24-hr food recall by the Research Coordinator to assess dietary intake for the day prior to hospital admission and at each follow-up time point. Will be analyzed for total kcal and macronutrient profile as covariates of treatment outcomes via Food Processor (ESHA, v. 11.1). In hospital, calories provided on meal trays and consumed are carefully tracked with calorie counts by Diet Technicians and nursing staff trained on protocol. These data provide proof of concept to demonstrate treatment fidelity and acceptability (formula intake to replace refused kcals). Administered at these timepoints: admission, 3 months, 6 months, 9 months, and 12 months.

Timing of Study Procedures:

YEAR	YR 1		YR 2		YR 3		YR 4		YR 5	
MO	1-6	7-12	1-6	7-12	1-6	7-12	1-6	7-12	1-6	7-12
Training	X	*	*	*	*	*	*	*		
DSMB meetings	X		X		X		X		X	
Lab Meetings	X	X	X	X	X	X	X	X	X	X
Number to be enrolled	7	10	10	10	10	10	10	7		
Number in follow-up		7	10	10	10	10	10	10	7	
Data	Collect/manage						Analyze		Write-up	

* Regular retraining to control inter-rater variability

3.2. Safety

Safety:

- A. **Medical Oversight:** This study has heavy medical oversight because it enrolls patients requiring hospitalization for medical instability. The treatment (ICR or HCR) is limited to the hospital stay; HCR is the current SOC at both clinical sites. All hospital procedures are consistent with SOC except for the questionnaires and REE measures. Routine electrolyte monitoring and replacement prescribed by physicians during the hospital stay will provide the safety data for AIM 2. Study blood samples in hospital will be added to the SOC blood draws performed by in-house phlebotomists and will not require additional venipuncture. At follow-up, blood draws will occur at the clinical laboratories staffed by pediatric phlebotomists in our respective outpatient clinics at UCSF and Stanford. Blood draws do carry a small additional risk, and indirect calorimetry can be uncomfortable, which will be clearly explained on the IRB-approved consent/assent forms. Study participants are followed openly while under a physician's care to ensure medical stability, however approximately 70% choose to receive that care from us. Many patients have a psychiatrist to manage psychiatric co-morbidities. If they do receive care elsewhere, that information is collected via our questionnaire, and they can continue in the study. Participants who become medically unstable during follow-up will be readmitted to hospital and receive SOC (not study diet).
- B. **Prospective monitoring of AEs:** AIM 2 specifies three electrolyte abnormalities that are monitored prospectively in all participants during hospitalization. All other AEs are tracked and recorded in OnCore, a clinical trial system.

- C. **Independent Data Safety Monitoring Board:** As a clinical trial, we have an independent DSMB per NICHD policy for clinical research monitoring. We have an expert on metabolic/hormonal biomarkers on our current DSMB and will convene the group prior to the start of enrollment, and each year of the clinical trial.

STATISTICAL ANALYSIS PLAN

Staged Analysis Plan and Reporting

This study includes pre-specified endpoints assessed at different times: (1) short-term outcomes during hospitalization (Aims 1 and 2) and (2) long-term clinical outcomes assessed over 12 months (Aim 3). The trial is powered for the primary short-term efficacy endpoint (Aim 1), and safety outcomes (Aim 2) are confined to the inpatient hospitalization period.

Analyses will be conducted in a staged manner. Stage 1 analyses will include all pre-specified short-term efficacy (Aim 1) and safety (Aim 2) outcomes and will be performed after completion of hospitalization data collection for all participants and database lock for these outcomes. Stage 2 analyses will include the pre-specified long-term primary endpoint (Aim 3) and will be conducted after completion of 12-month follow-up and final database lock.

This staged approach was pre-specified to allow timely evaluation and dissemination of short-term outcomes while preserving the integrity of long-term endpoint analyses. All endpoints, analytic methods, and timing of analyses were defined a priori. No data from long-term outcomes will be used to inform or modify Stage 1 analyses.

1. Aims and Objectives

Our study has three main aims.

We will compare:

AIM 1: Efficacy of ICR vs. HCR.

Efficacy: We hypothesize that ICR will restore medical stability faster than HCR during hospitalization, defined as an index, detailed below and comprised of the following: 1.) 24-hour HR ≥ 45 bpm, 2.) SBP ≥ 90 mmHg, 3.) temperature $\geq 35.6^{\circ}\text{C}$, 4.) orthostatic Δ HR ≤ 35 bpm, and 5.) orthostatic Δ SBP ≤ 20 mmHg.

AIM 2: Safety:

We hypothesize no evidence that ICR differs from HCR in incidence of:

i) hypophosphatemia (<3 mg/dL), ii) hypomagnesaemia (<1.8 mg/dL), and iii) hypokalemia (<3.5 mEq/L) during hospitalization

AIM 3: Clinical remission:

We hypothesize:

A. greater improvements in %mBMI and EDE-Q score over longitudinal follow-up (12-mo) in ICR than HCR; and

B. change in %mBMI will contribute to improvements in metabolic, hormonal, and psychological recovery markers (specified above); WS will modify these associations

2. Statistical Methods

2.1. Pool of Participants

- A. **Projected pool of eligible participants and accrual rate:** For StRONG, we approached N=301 eligible patients and enrolled ~3-4 participants/mo. for a total of N=120, including N=51 with AAN. For the proposed trial, we plan to enroll N=74 patients with AAN. Systems for the current trial are in place to allow us to begin enrollment immediately after our single IRB approval is in place. We will accrue ~2 participants/mo. over 3 yr until target enrollment is reached. We will retain 88% of participants over 12 mo. as we did in the parent trial.
- B. **Commitments of Site PIs, Research Teams, and Participants:** Clinical sites for the proposed trial are two major tertiary care children's hospitals with eating disorder inpatient programs attended by interdisciplinary care teams. Our large populations of youth up to age 24 are mostly students and keen to contribute to science. Participation is also facilitated by the fact that the investigators are also clinicians who are familiar and trusted faces in the units. Further, study treatments occur during medically indicated hospitalization, and most study procedures are SOC, therefore research adds minimal burden. Finally, we incentivize participation and follow-up with remuneration commensurate with the additional time commitment. PI Garber will be actively engaged in the daily trial activities, as they did to ensure the success of StRONG. Dr. Garber designed the proposed study and developed the dietary treatments. She will ensure that treatments are delivered with fidelity and train the staff (including kitchen and nursing) on treatments and data collection. Dr. Golden, a leading expert in hormonal and metabolic recovery in AN, published multiple studies examining the proposed biomarkers and indirect calorimetry in AN. He developed and trained the sites on these procedures. After his retirement, Dr. Kapphahn became the Stanford Site PI. Dr. Le Grange, the foremost expert in psychotherapy for eating disorders, oversaw all psychological aspects. After his retirement, Dr. Gorrell became the study psychologist. Dr. Cheng continues as faculty biostatistician and Dr. Gorrell serves as a lead analyst for the DCC. Together they ensure design, execution, and analysis without bias, and refine systems to manage, transfer and clean data. Drs. Buckelew and Downey (UCSF) and Kapphahn (Stanford) continue as study physicians, overseeing the medical care of study patients and mediating decisions about study protocol and clinical care (e.g., protocol deviations). Dr. Moscicki served as the safety liaison for the investigators, staff and DSMB.

2.2. Outcomes

Aim 1 Primary Efficacy Outcome:

Time to restore medical stability will be adjudicated by a five-point index: 1.) 24-hour HR $\geq$ 45 bpm, 2.) SBP $\geq$ 90 mmHg, 3.) temperature $\geq$ 35.6° C, 4.) orthostatic increase in HR $\leq$ 35 bpm, and 5.) orthostatic decrease in systolic BP $\leq$ 20 mmHg. Each criterion will be scored as "1" if met, "0" if unmet and missing (not scored) if not measured. Medical stability will be

considered restored when all the criteria are stable for 24 hours, allowing a maximum of one missing value (i.e., participants will be considered stable if meeting 4/4, 5/5 measured criteria). All randomized participants who receive at least one day of treatment, including those who withdraw at any time during the refeeding intervention, will be included in the modified intention to treat (mITT) analysis, excluding participants who withdraw without receiving any treatment. In the parent study most, but not all, participants met full stability criteria by hospital discharge (~85%), therefore we will plan to use survival analysis to allow right-censoring. As in the parent study, we expect randomization to balance distributions of baseline covariates, however we will confirm this with box plot, bar graph and descriptive statistics. If balanced, log rank test will compare time to achieve medical stability by arm while accounting for the correlation within sites. If unbalanced, Cox proportional hazard ratio model will be used to control for potential bias due to the unbalanced confounders. The assumption of proportional hazard ratio will be assessed by checking scaled Schoenfeld residuals against time and log-log survival curves.

Secondary analyses may consider

- 1) Subgroup analyses to assess if the treatment effect on short-term outcome time to medical stability is different across subgroups such as:
 - a. Age group, (e.g., 12-14, 15-18, >18 years old)
 - b. %mBMI at baseline: participants will be categorized into 2 groups by median of %mBMI and 4 groups by quantiles of %mBMI at baseline respectively. In addition to see if the treatment effect differs across the 2 and 4 baseline %mBMI groups, we will also specifically evaluate the treatment effect in the more extreme groups: below first quantile group and above the third quantile group.
 - c. Markers of WS (e.g., duration, magnitude, rate, energy balance) before admission will be considered for its moderating effect in both its original scale and categorized scale, categorized by median or quantiles.
 - d. Psychometric scores (e.g., eating disorder psychopathology, anxiety, mealtime distress) at baseline will be considered for their potential moderating effect in both their original scales and categorized scales, categorized by normal threshold, median or quantiles.
- 2) Combining the treatment groups if no significant results are found by treatment and identifying potentially important predictors of the short-term outcome using recently developed machine learning techniques for survival analysis. Potential predictors of the short-term outcome include age group, %mBMI at baseline, duration of illness before admission, % historical weight loss before admission, EDE-Q score at baseline, as discussed above.
- 3) Testing group differences on secondary short-term efficacy outcome including proportion achieving medical stability, change %mBMI vs. baseline, and time to normalize/restore HR.

Aim 2 Outcomes:

The occurrence of the AEs during hospitalization will be summarized with frequencies, percentages and 95% confidence intervals, and compared between treatment arms with Fisher's exact tests. These outcomes will be actively minimized by study physicians; we expect few AEs to occur. Other outcomes may include electrolyte nadirs and use of electrolyte supplementation.

Aim 3A Outcomes:

3A. Primary analysis for Aim 3 will longitudinally compare clinical remission components (over the span of 12-mo) in ICR compared to HCR, with (generalized) linear mixed effect models (GLMM for EDE-Q, LMM for %mBMI). Models will include time, treatment (ICR vs. HCR), and time×treatment interaction while controlling for any unbalanced baseline covariates. We will include study sites and patients as random effects to account for correlation within clusters. Assuming similar background effects/trends over time for both groups, the time×treatment interaction will provide us a mITT estimate for the effect of ICR vs. HCR on clinical remission parameters over time.

Secondary analyses will consider

- 1) subgroup analyses to assess if the treatment effects on longitudinal outcomes for EDE-Q scores and %mBMI are different across subgroups by including subgroup-treatment interactions in GLMM and LMM respectively. The subgroups we will consider include:
 - a. Age group, (e.g., 12-14, 15-18, >18 years old)
 - b. %mBMI at baseline: participants will be categorized into 2 groups by median of %mBMI and 4 groups by quantiles of %mBMI at baseline respectively. In addition to see if the treatment effect differs across the 2 and 4 baseline %mBMI groups, we will also specifically evaluate the treatment effect in the more extreme groups: below first quantile group and above the third quantile group.
 - c. duration of illness before admission will be considered for its moderating effect in both its original scale and categorized scale, categorized by median or quantiles.
 - d. Markers of WS (e.g., duration, magnitude, rate, energy balance) before admission will be considered for its moderating effect in both its original scale and categorized scale, categorized by median or quantiles.
 - e. Psychometric scores (e.g., eating disorder psychopathology, anxiety, mealtime distress) at baseline will be considered for their potential moderating effect in both their original scales and categorized scales, categorized by normal threshold, median or quantiles.
- 2) combining the treatment groups if no significant results are found by treatment and identifying potentially important predictors of longitudinal outcomes using dominance analysis or random forest or elastic net, as appropriate.

- a. Potential baseline predictors of the long-term outcomes include age group, %mBMI at baseline, duration of illness before admission, % historical weight loss before admission, EDE-Q score at baseline, as discussed above.
 - b. Potential predictors at discharge (intermediate variables) of longitudinal outcomes include %mBMI at discharge, length of hospital stay, number of readmissions to hospital. If any intermediate variable is identified as a significant predictor of later longitudinal outcomes, we will further examine if the treatment influences the intermediate variable and subsequently influences the protracted outcomes using a mediation analysis.
- 3) similar analyses on secondary longitudinal outcomes (e.g., including proportion achieving EDE-Q scores within range of population norms; time to normalize menses)

Aim 3B Outcomes:

LMM and linear mixed-effect lag models (LMLM) will respectively evaluate the impact of Δ %mBMI on contemporaneous and subsequent changes in metabolic (REE, T3), hormonal [ROM, LH, FSH, E2 (girls); testosterone (boys); leptin; cortisol], and psychological recovery (EDE-Q, BI-AAQ, WBIS-M, SATAQ-4R) during hospitalization and the longitudinal follow-up period. LMM will have recovery markers as outcomes, include time, %mBMI, and time $\times$ %mBMI interaction as fixed effects, with site and patients as random effects. LMLM will replace concurrent %mBMI with last %mBMI for its effect on the subsequent recovery markers. If the relation between %mBMI and recovery markers varies over time, both the main effect of %mBMI and time $\times$ %mBMI interaction will be used to assess the %mBMI-recovery markers relation over time. Additionally, the total change of %mBMI over 12 mo, a measure for the cumulative weight change over the full follow-up time period, will be calculated and assessed for its association with metabolic, hormonal, and psychological recovery outcomes at 12 mo. Secondary analyses will explore WS and other baseline characteristics (e.g., pubertal stage, psychiatric medication [yes/no]) as potential moderators. Since ~90% will be female, we do not expect meaningful sex differences. To further understand whether weight gain is sufficient, we will assess the shape and magnitude of the Δ %mBMI and recovery markers over time. A significant positive relation will indicate a positive impact of weight gain on recovery; a plateau in biomarkers within published pediatric reference ranges or published norms for psychological measures will indicate that sufficient %mBMI was achieved. Additionally, to examine the mechanism of treatment on recovery, mediation analyses will consider direct and indirect effects of treatment around and through weight gain on recovery markers. The step-down procedure will be used to adjust for group comparisons on the multiple outcomes.

2.3. Power and Sample-Size Considerations

Aim 1: This study is powered to detect a significant difference in time to medical stability (within hospital) between ICR and HCR, based on findings from the parent trial. This aim is executed in hospital and therefore we do not experience attrition [patients cannot leave unless against medical advice with signed legal documentation by parent(s).] However, in the parent trial, 6% of randomized participants did withdraw prior to receiving treatment. Treatment effect for the proposed test group was estimated from StRONG participants with

AN treated with HCR (mean 49 kcal/kg), approximating ICR, vs. AAN. N=34 per arm would provide 80% power on a 2-sided 0.05-level test to detect the expected difference in time to medical stability between ICR vs. HCR of 3.85 days (95% CI 1.09, 6.62). We will plan for a total sample of N=74 to allow for withdrawals and loss to follow-up (see Aim 3).

Aim 2: Outcomes for this aim are monitored closely as AEs and actively minimized through careful medical monitoring and protocolized correction (oral electrolyte replacement). [This aim is not powered.]

Aim 3A: Given the estimates from StRONG, 0.65 intraclass correlation of %mBMI at discharge, 3, 6, and 12 mo, standard deviation of 12%, and 88% retention, a starting N=74 over 4 follow-ups will provide 80% power in two-sided tests with a type I error level of 5% to detect a 7.6% group difference in %mBMI. For context as to how this may relate to the proposed recovery markers, in StRONG the difference in %mBMI in AAN participants with and without ROM was 102.6 ± 12.2 vs. 95.0 ± 3.4 %; in a larger study, PI Golden found 106.1 ± 11.7 vs. 94.2 ± 8.9 .

2.4. Data Management

- a. **DCC:** Drs. Cheng and Gorrell at UCSF lead the DCC, autonomous and independent of the clinical sites. It is housed in the UCSF Department of Preventive and Restorative Dental Sciences (PRDS). The department is staffed primarily by statisticians who conduct data-intensive research and multicenter trials. They are equipped with an independent computer network with high-level security for protected health information. The network is maintained by in-house IT, including ongoing maintenance, installation, software upgrades and structural components such as cabling and servers. As lead biostatistician, Dr. Cheng oversees the data analysis and guides the analytic design. She led the DCC for StRONG and therefore is familiar with all of the data systems, study design, analytic plans, and maintenance of autonomous but close working relationship with clinical sites. Dr. Gorrell oversees data storage, integrity, and monitoring.
- b. **Electronic data capture:** Both clinical sites use EPIC EMR for clinical data in hospital and outpatient clinics. Computer code shared between sites allows automatic download of study data from the EMR. Analytics- reporting technology managers wrote the code in SQL, which is wrapped in Crystal Reporting software and presented via EPIC-Crystal integration. Daily reports from EPIC are sent to the DCC through Secure Box, where they are encrypted on a secure remote server. Inpatient nutrition data are downloaded from the foodservice management software Computrition and are similarly transferred. The secure enterprise system for clinical research management is the Online Collaborative Research Environment (OnCore), which is accessed online for all aspects of study management, including randomization, data collection and management, participant demographics, participant tracking, AE reporting and study timelines. Validated questionnaire scoring algorithms are programmed into Qualtrics; scores are exported for analysis using the Data Export Utility.
- c. **Confidentiality:** Loss of confidentiality is a recognized risk of participating in clinical research since protected health information, medical history, and demographics are used for the study. Loss of privacy may lead to problems with

insurability or social stigmatization. We have minimized this risk by utilizing existing secure systems (e.g., EMR) to ensure confidentiality. After extraction from the EMR, data are de-identified and handled by ID number only. No publications will include the names or identifying information about study participants.

2.5. Retention and Attrition

The design and timeline of the proposed trial are similar to StRONG, in which we successfully met our enrollment target and retained 88% of participants with high-quality (few missing) data and no safety concerns. We have accomplished this through enormous team effort, including regular investigator and staff meetings, ongoing training with clinical/nursing staff, and cross-training between clinical sites to minimize site differences. Finally, our prior work has been peer-reviewed through high-impact journals, presentations at international conferences, and participation in a national consortium.

The proposed iStRONG trial will utilize the StRONG infrastructure to expand the scope of our research into areas in which the study PIs have experience and both sites have equipment and personnel to support.

APPENDIX

1. Amendments to the Protocol

Appendix Table 1. Amendments to the Protocol

Protocol version	Section changed	Amendment description & reason	IRB approval and/or other documentation date
1.1	3.2: Data Collection	The MINI-KIDD was reduced to modules K and L only to reflect substance use and alcohol modules. Indirect handheld calorimetry instead of ambulatory metabolic cart for less patient burden. Stool sample collection at UCSF site only. Quick daily in-patient check-in surveys for SUDS and PANAS-C at UCSF site only.	IRB approved 3/11/2021
1.2	3.2: Data Collection	Updated blood volumes collected at admission and discharge. Changes to psychometric self-report questionnaires: removed PHQ-A for PHQ-8, Removed MINI-KIDD modules L and K for CRAFFT. Added GAD-7, DTS, EDE-15-Y, PAQ, AARP, SSS, PVA, Treatment Preference & Perception, SUDS, removed PANAS-C. These surveys are only for secondary aims for the study.	IRB approved 09/09/2021
1.3	2.1: Study Population	Inclusion criteria were clarified to reflect our HR inclusion criteria for the study. Nighttime heart <45bpm and/or daytime heart rate <50BPM. This was discussed with the Data and Statistical Monitoring Board	IRB approved 08/12/2024

Protocol version	Section changed	Amendment description & reason	IRB approval and/or other documentation date
		at study startup but was mistakenly left out of IRB approved protocol.	
1.4	3.1: Key Personnel	Upon Dr. Neville Golden's retirement, there was a discontinuation of the MPI agreement between UCSF and Stanford for the iStRONG trial. Dr. Cynthia Kapphahn was appointed as the new Site PI for Stanford for the trial's completion and Dr. Andrea Garber will serve as the sole PI for the study. Additionally, Dr. Megan Vo was added as the key personnel to support with the trial in its final year.	IRB approved 12/10/2024

2. Amendments to the Statistical Analysis Plan

Appendix Table 2. Amendments to the Statistical Analysis Plan.

Section changed	Amendment description & rationale	Amendment date & documentation
	Amendment 1:	
	Rationale:	