Clinical Islet Transplantation Consortium Protocol 07

STATISTICAL ANALYSIS PLAN

Principal Investigators
Bernhard Hering, MD
Xunrong Luo, MD, PhD
Olle Korsgren, MD, PhD
Nicole Turgeon, MD
Ali Naji, MD, PhD
Andrew Posselt, MD, PhD
Camillo Ricordi, MD
James Shapiro, MD, PhD
Dixon Kaufman, MD, PhD

Medical Monitors
Nancy Bridges, MD
Thomas L. Eggerman MD, PhD

Study Statisticians
Eric D Foster, PhD
Jon Yankey, MS

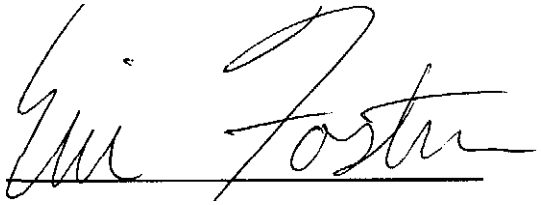
Senior Regulatory Officer
Julia Goldstein, MD

Project Manager
Allison Priore, BS

Supported By:
The National Institute of Allergy and Infectious Diseases (NIAID)
The National Institute of Diabetes & Digestive & Kidney Diseases (NIDDK)

VERSION 3.0
August 21, 2015

I have read and accept the changes that have been made to this SAP since Version 2.0.



Eric D Foster, PhD
Department of Biostatistics
Clinical Trials Statistical and Data Management Center
University of Iowa
2400 UCC
Iowa City, Iowa 52242
Phone: 319-384-4188
e-mail: eric-foster@uiowa.edu

Justification of changes since Version 2.0:
There are two issues with the CIT-07 secondary Bayesian subgroup analysis that this change to the statistical analysis plan (SAP) seeks to address. First, the current SAP makes no provisions for those centers that, upon combining subjects across multiple CIT protocols, cannot reach the targeted 12 subject number. Combining subjects across multiple CIT protocols, even in those instances where the 12 subject threshold cannot be met, would lead to larger individual center power and global power. Second, for some centers, it may be possible to reach the 12 subject threshold in multiple ways. In such cases, subject selection may be perceived as subjective.

The proposed modification is to perform two analyses. The first analysis will combine all subjects recruited at a CIT-07 center that participated in any of the CIT-02, -03, -04, -05, -06, and -07 protocols. In this manner, the analysis will incorporate the maximum possible number of CIT transplant recipients at each center, regardless of whether the sum totals less than or greater than 12. Moreover, the decision to include all subjects removes any perception of subjectivity with regard to subject selection for the analysis.

The second analysis will perform the same secondary Bayesian subgroup analysis on the CIT-07 data alone. This analysis should provide insight into how the first analysis' inclusion of subjects from other protocols impacts the CIT-07 alone results.

Changes made since SAP Version 2.0:

Section/Page#	SAP Text Original	SAP Text Change (bold)
3.7.2.1.3 Page 23	***3.7.2.1.3 Estimation of rate at individual centers*** Twelve subjects at one center is considered sufficient to demonstrate comparable efficacy at that center. This is based on the power calculations given in SAP Appendix 8.1.	***3.7.2.1.3 Estimation of rate at individual centers*** Twelve subjects at one center is considered sufficient to demonstrate comparable efficacy at that center. This is based on the power calculations given in SAP Appendix 8.1. **Two methods for the estimation of the rate at individual centers are used: 1) estimation based on the complete set of endpoint data from CIT-07 centers using data from CIT protocols 02, 03, 04, 05, 06, and 07 and 2) estimation based on the endpoint data from CIT-07 centers using data from CIT-07 alone.**
3.7.2.1.3.1 Page 23	**3.7.2.1.3.1 For centers enrolling at least 12 subjects in CIT-07** To evaluate an individual center, that center must either (1) enroll at least 12 subjects in CIT-07 or (2) enroll a combined total of at least 12 subjects in CIT-07 and one of the site-specific Phase 2 studies (CIT-02, 03, 04, or 05). If a center enrolls at least 12 subjects in CIT-07, the criteria for evaluation of that center require that the data from CIT-07 satisfy both (a) and (b) below: a) The overall primary result of CIT-07 is positive: that is, the primary analysis, which constructs a one-sided 95% confidence interval for the overall probability favorable outcome assuming no between center variability, rules out favorable outcome rate of 0.50 or less.	**3.7.2.1.3.1 ~~For centers enrolling at least 12 subjects in CIT-07~~ Estimation of the rate at individual centers based on the complete set of endpoint data from CIT-07 centers using data from CIT protocols 02, 03, 04, 05, 06, and 07** ~~To evaluate an individual center, that center must either (1) enroll at least 12 subjects in CIT-07 or (2) enroll a combined total of at least 12 subjects in CIT-07 and one of the site-specific Phase 2 studies (CIT-02, 03, 04, or 05).~~ ~~If a center enrolls at least 12 subjects in CIT-07, the criteria for evaluation of that center require that the data from CIT-07 satisfy both (a) and (b) below:~~ **Those CIT-02, -03, -04, -05, -06, and -07 subjects enrolled at a CIT-07 participating center are assessed for meeting the CIT-07 primary endpoint definition. If a center enrolls at least 12 subjects from the combination of the CIT-02, -03, -04, -05, -06, and -07 protocols, the criteria**

Section/Page#	SAP Text Original	SAP Text Change (bold)
		for evaluations of that center require that the resulting analysis satisfy both (a) and (b) below: a) The overall primary result of **the combined protocols** is positive: that is, the primary analysis, which constructs a one-sided 95% confidence interval for the overall probability **of** favorable outcome assuming no between center variability, rules out **a** favorable outcome rate of 0.50 or less.

Section/Page#	SAP Text Original	SAP Text Change (bold)
3.7.2.1.3.2 Page 24	**3.7.2.1.3.2 For centers enrolling a combined total of at least 12 subjects in CIT-07 and one site-specific phase 2 study (CIT-02, 03, 04, or 05)** In this case the entire data set for study CIT-07 will be augmented with the data from the site-specific phase 2 study at the center of interest. The subgroup analysis described in SAP section 3.7.2.1.1 will be fit to the augmented data set. The lower 90% probability bound for the center of interest will be examined and compared to 0.45. If the lower 90% probability bound is larger than 0.45 then that center's performance will be considered comparable to that of the study as a whole.	~~**3.7.2.1.3.2 For centers enrolling a combined total of at least 12 subjects in CIT-07 and one site-specific phase 2 study (CIT-02, 03, 04, or 05)**~~ ~~In this case the entire data set for study CIT-07 will be augmented with the data from the site-specific phase 2 study at the center of interest. The subgroup analysis described in SAP section 3.7.2.1.1 will be fit to the augmented data set. The lower 90% probability bound for the center of interest will be examined and compared to 0.45. If the lower 90% probability bound is larger than 0.45 then that center's performance will be considered comparable to that of the study as a whole.~~ **3.7.2.1.3.2 Estimation of the rate at individual centers based on the endpoint data from CIT-07 centers using data from CIT-07 alone** **The subgroup analysis described in SAP section 3.7.2.1.1 will also be fit to a dataset comprised of the CIT-07 subjects alone. If a center enrolls at least 12 subjects in CIT-07, the criteria for evaluations of that center require that the resulting analysis satisfy both (a) and (b) below:** **a) The overall primary result of CIT-07 is positive: that is, the primary analysis, which constructs a one-sided 95% confidence interval for the overall probability of favorable outcome assuming no between center variability, rules out a favorable outcome rate of 0.50 or less.** **b) The favorable outcome rate at that center, under the Bayesian model, is estimated to be at least 0.45 with posterior probability 0.90. That is, the lower 90% probability bound is at least 0.45.**

Section/Page#	SAP Text Original	SAP Text Change (bold)
3.7.2.1.4 Page 24	**3.7.2.1.4 Criteria** Deeming that performance of an individual center is comparable to that of the study as a whole will also require that criterion (a) above is met, that is, a favorable outcome rate of 0.50 or less for the study CIT-07 as a whole is ruled out by a frequentist analysis and a 95% one sided confidence interval. In addition, criterion (b) is met for that center.	~~**3.7.2.1.4 Criteria**~~ ~~Deeming that performance of an individual center is comparable to that of the study as a whole will also require that criterion (a) above is met, that is, a favorable outcome rate of 0.50 or less for the study CIT-07 as a whole is ruled out by a frequentist analysis and a 95% one sided confidence interval. In addition, criterion (b) is met for that center.~~
3.7.2.1.4.1 Page 24	3.7.2.1.4.1 **Power calculations (α=2, β=1.5)**	3.7.2.1.4.~~1~~ **Power calculations (α=2, β=1.5)**
Page 64	**Proposed Evaluation of Centers** **For Centers Enrolling at least 12 subjects in CIT-07** The proposed criteria require that a center must enroll either at least 12 subjects in CIT- 07 or a total of at least 12 in CIT-07 and one of studies CIT-02, 03, 04 and 05 combined. (Each center will participate in CIT-07 and also one of CIT-02, 03, 04 and 05). If a center enrolls at least 12 subjects on CIT-07 the criteria for the center require that the Bayesian analysis of the data from CIT-07 is such that both: a)The overall primary result of CIT-07 is positive: that is the primary analysis, which constructs a one sided 95% confidence interval for the overall probability favorable outcome assuming not between center variability, rules out favorable outcome rate of 0.50 or less.	**Proposed Evaluation of Centers** ~~**For Centers Enrolling at least 12 subjects in CIT-07**~~ ~~The proposed criteria require that a center must enroll either at least 12 subjects in CIT- 07 or a total of at least 12 in CIT-07 and one of studies CIT-02, 03, 04 and 05 combined. (Each center will participate in CIT-07 and also one of CIT-02, 03, 04 and 05).~~ ~~If a center enrolls at least 12 subjects on CIT-07 the criteria for the center require that the Bayesian analysis of the data from CIT-07 is such that both:~~ **Twelve subjects at one center is considered sufficient to demonstrate comparable efficacy at that center. This is based on the power calculations given in this Appendix.** **Two methods for the estimation of the rate at individual centers are used: 1) estimation based on the complete set of endpoint data from CIT-07 centers using data from CIT protocols 02, 03, 04, 05, 06, and 07 and 2) estimation based**

Section/Page#	SAP Text Original	SAP Text Change (bold)
		on the endpoint data from CIT-07 centers using data from CIT-07 alone. **Estimation of the rate at individual centers based on the complete set of endpoint data from CIT-07 centers using data from CIT protocols 02, 03, 04, 05, 06, and 07** **Those CIT-02, -03, -04, -05, -06, and -07 subjects enrolled at a CIT-07 participating center are assessed for meeting the CIT-07 primary endpoint definition. If a center enrolls at least 12 subjects from the combination of the CIT-02, -03, -04, -05, -06, and -07 protocols, the criteria for evaluations of that center require that the resulting analysis satisfy both (a) and (b) below:** a)The overall primary result of **the combined protocols** ~~CIT-07~~ is positive: that is**,** the primary analysis, which constructs a one sided 95% confidence interval for the overall probability **of** favorable outcome assuming no between center variability, rules out **a** favorable outcome rate of 0.50 or less**.**
Page 65	**For Centers enrolling at least 12 subjects in CIT-07 and one of CIT-02, 03, 04, or 05 combined** Should Center A enroll at least 6 subjects in CIT-07 and a total of at least 12 subjects in CIT-07 and one of 02, 03, 04 or 05, the following analysis will be done. The entire data set for study CIT-07 will be augmented with the data from Center A from the relevant CIT pilot study (the pilot study in which Center A participates). The subgroup analysis of section 3.7.2.1 will be run on the augmented data set (the augmented data set is anticipated to include the 48 subjects in CIT-07 plus all subjects enrolled in the pilot study at Center A). The	~~**For Centers enrolling at least 12 subjects in CIT-07 and one of CIT-02, 03, 04, or 05 combined**~~ ~~Should Center A enroll at least 6 subjects in CIT-07 and a total of at least 12 subjects in CIT-07 and one of 02, 03, 04 or 05, the following analysis will be done. The entire data set for study CIT-07 will be augmented with the data from Center A from the relevant CIT pilot study (the pilot study in which Center A participates). The subgroup analysis of section 3.7.2.1 will be run on the augmented data set (the augmented data set is anticipated to include the 48 subjects in CIT-07 plus all subjects enrolled in the pilot study at Center A). The~~

Section/Page#	SAP Text Original	SAP Text Change (bold)
	lower 90% probability bound for center A will be examined and compared to 0.45. If bound is larger than 0.45 then criterion (b) is met, and if criterion (a) is also met for the CIT-07 data set, then center A will be considered consistent with the results of CIT-07 as a whole.	~~lower 90% probability bound for center A will be examined and compared to 0.45. If bound is larger than 0.45 then criterion (b) is met, and if criterion (a)is also met for the CIT-07 data set, then center A will be considered consistent with the results of CIT-07 as a whole.~~ **Estimation of the rate at individual centers based on the endpoint data from CIT-07 centers using data from CIT-07 alone** **The subgroup analysis described in SAP section 3.7.2.1.1 will also be fit to a dataset comprised of the CIT-07 subjects alone. If a center enrolls at least 12 subjects in CIT-07, the criteria for evaluations of that center require that the resulting analysis satisfy both (a) and (b) below:** **a) The overall primary result of CIT-07 is positive: that is, the primary analysis, which constructs a one-sided 95% confidence interval for the overall probability of favorable outcome assuming no between center variability, rules out a favorable outcome rate of 0.50 or less.** **b) The favorable outcome rate at that center, under the Bayesian model, is estimated to be at least 0.45 with posterior probability 0.90. That is, the lower 90% probability bound is at least 0.45.**

Clinical Islet Transplantation Consortium

Islet Transplantation in Type 1 Diabetes

Statistical Analysis Plan for CIT-07

Table of Contents

Glossary of Abbreviations

ACE	American College of Endocrinology
AE	adverse event
AIR_{glu}	acute insulin response to glucose
ATG	rabbit anti-thymocyte globulin
BMI	body mass index
BG	blood glucose
BW	body weight
CGMS	Continuous Glucose Monitoring System®
CIT	Clinical Islet Transplantation
DDS	Diabetes Distress Scale
DI	disposition index
DSMB	Data Safety Monitoring Board
EQ-5D	European Quality of Life
FSIGT	frequently-sampled intravenous glucose tolerance
GFR	glomerular filtration rate
HbA1c	glycosylated hemoglobin
HFS	Hypoglycemia Fear Scale
HLA	histocompatability antigen
HSA	human serum albumin
IEQ	islet equivalents
ITT	intent to treat

LI	lability index
MAGE	mean amplitude of glycemic excursions
MCMC	Monte Carlo Markov Chain
MMTT	mixed-meal tolerance test
NCI	National Cancer Institute
NIH	National Institutes of Health
OHS	Overall Health Status
PAID	Problem Areas in Diabetes
QOL	quality of life
SAE	serious adverse event
SAP	statistical analysis plan
SI	insulin sensitivity
SMC	Summary Mental Component
SPC	Summary Physical Component
T1D	Type 1 Diabetes
TCAE	Terminology Criteria for Adverse Events
ULN	upper limit of normal

1. Introduction

Type 1 diabetes (T1D) afflicts nearly 2 million people in the United States, most of them children or young adults. Exogenous insulin, administered by multiple injections or by a continuous subcutaneous infusion from a wearable pump, allows long term survival in those who develop the disease, and most who are treated in this way will have a very good health-related quality of life (QOL). However, insulin therapy does not provide normal glycemic control, and long-term survivors commonly develop vascular complications such as diabetic retinopathy (the most common cause of adult blindness) and diabetic nephropathy (the most common indication for adult kidney transplantation). A small minority of individuals with T1D develop hypoglycemia unawareness, a condition that is life threatening, is associated with severe deterioration in QOL and activity restriction, and is not amenable to medical therapy.

The hope of achieving near-normal glucose control without hypoglycemia in T1D has provided the impetus for developing effective strategies for β-cell replacement via pancreas or isolated islet transplantation. Islet transplantation is accomplished by a procedure in which the islets are infused into the portal vein. While this procedure is not without risk, the procedural morbidity is much less than that of whole pancreas transplant.

While about 80% of whole pancreas transplant recipients will be insulin independent at one year after their transplant, less than 10% of 447 islet recipients transplanted between 1990 and 1999 achieved one year insulin independence. This was attributed to low engrafted islet mass combined with high metabolic demand imposed by glucocorticoids used to prevent rejection. In the year 2000, the group from Edmonton reported a series of 7 consecutive islet transplant recipients treated with islets from multiple donors and a glucocorticoid-free immunosuppressive regimen [1]. These islet recipients were insulin free at follow-up ranging from 4.5 to 15 months. All of the recipients had experienced severe hypoglycemic episodes prior to transplant, and afterwards, none did. The efficacy of the Edmonton approach has now been confirmed by several other centers, and represents a major breakthrough in the field.

A Clinical Islet Transplantation (CIT) consortium has been formed to conduct a multicenter trial with the goal of providing strong scientific evidence that the rate of favorable outcome in transplanted subjects is high enough to justify the risks of the islet transplant procedure and the required immunosuppression.

This document briefly describes the study design and provides a detailed statistical analysis plan (SAP) for treatment efficacy and safety assessments. Details of all study aspects are given in the formal study protocol (CIT-07).

2. Study Design and Objectives

2.1 *Study Design*

This is a prospective, single-arm, multi-center study in islet transplantation. The centers participating in this phase 3 study will also undertake separate, phase 2 studies in islet transplantation, using innovative manufacturing and/or immunosuppressive regimens. In order to avoid bias in selection of subjects for these studies, eligible subjects will be randomized, prior to transplantation, to participate either in the phase 3 or a site-specific phase 2 study. The participating centers will treat a total of 48 study subjects accrued over 24 months in this phase 3 study.

2.2 *Study Objective*

2.2.1 Primary Objective

The primary objective is to demonstrate, in a multi-center, single-arm study, the safety and efficacy of islet transplantation for the treatment of T1D in subjects with hypoglycemia unawareness and a history of severe hypoglycemic episodes.

2.2.2 Selection of Subjects

Please see Section 3 of the CIT-07 protocol for subject inclusion and exclusion criteria.

2.2.3 Study Treatment Regimen

2.2.3.1 Investigational Agent: Allogeneic Islets

The investigational agent is the Purified Human Pancreatic Islet product. The final product is a sterile suspension of ≥70% viable, ≥30% pure, allogeneic human purified islets in 200 mL of transplant media containing 2.5% human serum albumin (HSA), 25 mM Hepes for administration by intraportal infusion. Each product lot may comprise up to 3 bags containing 200 mL each. The final product dose is ≥5,000 islet equivalents (IEQ)/kg recipient body weight (BW) for the first infusion, and ≥4,000 IEQ/kg recipient BW for subsequent infusions.

The transplant procedure involves infusion of the final product into a branch of the portal vein, which is accessed by percutaneous transhepatic cannulation using ultrasound guidance and fluoroscopic localization of the liver, or under direct visualization via a minilaparotomy.

2.2.3.2 Immunosuppression

For the initial islet transplant, the study medication is administered concomitantly with a consensus regimen of immunosuppressive and anti-inflammatory medications that includes rabbit anti-thymocyte globulin (ATG, Thymoglobulin®), sirolimus (Rapamune®), tacrolimus (Prograf®) and etanercept (Enbrel®). For subsequent allogeneic islet transplants, daclizumab (Zenapax®) will be used instead of Thymoglobulin®.

2.3 *Study Endpoints*

2.3.1 Primary Endpoint

The primary endpoint for this study is the proportion of subjects with an HbA1c <7.0% at Day 365 AND free of severe hypoglycemic events from Day 28 to Day 365 inclusive following the first islet transplant, with the day of transplant designated Day 0. The primary aim of the analysis is to estimate the true rate of this outcome in subjects analyzed by intention-to-treat (ITT) as defined in SAP section 3.2.1.

2.3.2 Secondary Endpoints

Because there are a large number of secondary endpoints, it is impractical to account for all multiple comparisons. However, a few secondary endpoints have been identified as **key** secondary endpoints.

2.3.2.1 Key secondary endpoints

The target level for HbA1c chosen for this study is 7.0%. This value was chosen because it is the level recommended by the American Diabetes Association and is considered to be the clinically relevant goal for subjects with T1D. A HbA1c level of 6.5% is the goal recommended by the American College of Endocrinology (ACE). We have included achieving a HbA1c level of 6.5%, alone and as a composite with freedom from severe hypoglycemic events at 1 year after the first islet transplant, as **key** secondary endpoints because they correspond to the ACE recommendations and will be of interest to the medical community.

The key secondary endpoints are the following:

1) The proportion of subjects with an HbA1c <7.0% AND free of severe hypoglycemic events from Day 28 to Day 730, inclusive, after the first islet transplant.
2) The proportion of subjects with HbA1c ≤ 6.5% at one year after the first islet transplant AND free of severe hypoglycemic events from Day 28 to Day 365 after the first islet transplant.
3) The proportion of subjects with HbA1c ≤ 6.5% at two years after the first islet transplant AND free of severe hypoglycemic events from Day 28 to Day 730 after the first islet transplant.
4) The proportion of subjects free of severe hypoglycemic events from Day 28 to Day 365 after the first islet transplant.
5) The proportion of subjects free of severe hypoglycemic events from Day 28 to Day 730 after the first islet transplant.
6) The proportion of subjects with HbA1c <7.0% at one year after the first islet transplant.

7) The proportion of subjects with HbA1c <7.0% at two years after the first islet transplant.
8) The proportion of subjects with HbA1c ≤6.5% at one year after the first islet transplant.
9) The proportion of subjects with HbA1c ≤6.5% at two years after the first islet transplant.
10) The proportion of insulin-independent subjects at one year after the first islet transplant.
11) The proportion of insulin-independent subjects at two years after the first islet transplant.

2.3.2.2 Additional efficacy endpoints

Other secondary efficacy endpoints include the following:

- The proportion of subjects with an HbA1c <7.0% AND free of severe hypoglycemic events from Day 28 to Day 730, inclusive, after the final islet transplant.
- The proportion of subjects with HbA1c ≤ 6.5% at one year after the final islet transplant AND free of severe hypoglycemic events from Day 28 to Day 365 after the final islet transplant.
- The proportion of subjects with HbA1c ≤ 6.5% at two years after the final islet transplant AND free of severe hypoglycemic events from Day 28 to Day 730 after the final islet transplant.
- The proportion of subjects free of severe hypoglycemic events from Day 28 to Day 365 after the final islet transplant.
- The proportion of subjects free of severe hypoglycemic events from Day 28 to Day 730 after the final islet transplant.
- The proportion of subjects with HbA1c <7.0% at one year after the final islet transplant.
- The proportion of subjects with HbA1c <7.0% at two years after the final islet transplant.
- The proportion of subjects with HbA1c ≤6.5% at one year after the final islet transplant.
- The proportion of subjects with HbA1c ≤6.5% at two years after the final islet transplant.

- The proportion of insulin-independent subjects at one year after the final islet transplant.
- The proportion of insulin-independent subjects at two years after the final; islet transplant

At 75 ± 5 days following the first and subsequent transplant(s):

- The percent reduction in insulin requirements
- HbA1c
- Mean amplitude of glycemic excursions (MAGE) [2]
- Glycemic lability index (LI) [3]
- Ryan hypoglycemia severity (HYPO) score [3]
- Basal (fasting) and 90-min glucose and C-peptide derived from the mixed-meal tolerance test (MMTT)
- β-score [4]
- C-peptide/(glucose·creatinine) ratio
- Acute insulin response to glucose (AIR_{glu}), insulin sensitivity (SI), and disposition index (DI) derived from the insulin-modified frequently-sampled intravenous glucose tolerance (FSIGT) test [5]
- Glucose variability [6] and hypoglycemia duration [7] derived from the continuous glucose monitoring system® (CGMS)
- QOL measures, as defined in SAP section 3.6.5.

At 365 ± 14 days following the first and final islet transplant(s):

- The percent reduction in insulin requirement
- HbA1c
- MAGE
- LI
- Clarke score [14]
- HYPO score
- Basal (fasting) and 90-min glucose and C-peptide (MMTT)
- β-score
- C-peptide/(glucose·creatinine) ratio
- AIR_{glu}, SI, and DI derived from the FSIGT test [5]
- QOL, as defined in SAP section 3.6.5.
- The proportion of subjects receiving a second islet transplant
- The proportion of subjects receiving a third islet transplant
- Rate of favorable outcome at each center preparing islets (rate of subjects with an HbA1c <7.0% and free of severe hypoglycemic events)

At two years following the final islet transplant:

- The percent change from baseline insulin requirements.
- The number of severe hypoglycemic events.
- HbA1c.
- Clarke score.
- Basal (fasting) and 90-min glucose and c-peptide (MMTT).
- β-score.
- C-peptide: (glucose• creatinine) ratio.
- CGMS.
- QOL.

2.3.2.3 Safety endpoints

At 75 ± 5 days following each transplant and 365 ± 14 days following the first and final islet transplant(s):

- The incidence and severity of adverse events (AEs) related to the islet transplant procedure, including: bleeding (>2 g/dL decrease in hemoglobin concentration); segmental portal vein thrombosis; biliary puncture; wound complication (infection or subsequent hernia); and increased transaminase levels >5 times upper limit of normal (ULN)
- The incidence and severity of AEs related to the immunosuppression including: allergy; reduction in glomerular filtration rate (GFR); increase in urinary albumin excretion; addition or intensification of anti-hypertensive therapy; addition or intensification of anti-hyperlipidemic therapy; oral ulcers; lower extremity edema; gastrointestinal toxicity; neutropenia, anemia, or thrombocytopenia; viral, bacterial, or fungal infections; and benign or malignant neoplasms
- The incidence of change in the immunosuppression drug regimen
- The incidence of immune sensitization defined by presence of anti-HLA (histocompatibility antigen) antibodies absent prior to transplantation

At 365 ± 14 days following the first islet transplant:

- The incidence of worsening retinopathy as assessed by change in retinal photography

2.4 *Sample Size and Power Calculations*

The primary outcome (favorable outcome) is defined as the dichotomous event HbA1c <7.0% at Day 365 AND free of severe hypoglycemic events from Day 28 to Day 365, inclusive, following the first islet transplant, with Day 0 designated as the day of transplant. The sample size assessment and power calculations are based on the rate of this favorable outcome.

Historical data were available on a total 131 islet alone transplants performed at the University of Alberta, the University of Miami, the University of Minnesota, the University of Pennsylvania, Emory University and Northwestern University A summary of the observed favorable outcome rates are presented in the following table.

Table 1 - Favorable outcomes in the six participating clinical centers

Center	Number of Subjects Transplanted	Number with favorable outcome	Percent with favorable outcome
University of Alberta	68	49	72
University of Miami	21	17	81
University of Minnesota	20	16	80
University of Pennsylvania	9	6	75
Emory University*	8	7	88
Northwestern University*	5	4	80
Total	131	99	76

*co-investigator institutions

The observed favorable outcome rate is higher than 70% for all centers. To determine sample size, we constructed exact one-sided 95% confidence intervals [8] for selected observed favorable outcome rates. These lower bounds are listed in Table 2. For a selected observed rate, we can be 95% confident that the true (but unobservable) favorable outcome rate is at least as large as the lower bound of the confidence interval. With 95% confidence, any rate less than the tabulated endpoint will be excluded as a potential value for the true favorable outcome rate.

Table 2 - The highest favorable outcome to be ruled out with 95% confidence

	Sample Size			
	24	36	48	60
Observed Favorable Outcome Rate	Lower Confidence Bounds for Exact One-sided Binomial Confidence Intervals For Favorable Outcome Rate (%)			
10%	1.5	3.9	4.2	4.4
20%	8.6	9.5	11.8	12.0
30%	14.6	18.2	18.6	20.4
40%	24.6	25.3	27.7	29.3
50%	31.9	35.3	37.4	38.7
60%	39.7	46.0	47.5	48.6
70%	52.1	54.5	58.2	58.8
80%	61.1	66.6	67.2	69.6
90%	76.0	76.3	79.3	81.2

The best clinical judgment of the investigators is that a 50% or larger favorable outcome rate would be clinically meaningful. Based on past experience we would expect to

observe a rate larger than 70%. With a sample size of 48 transplanted subjects, 31 subjects (65%) would need to achieve a favorable outcome for the exact 95% lower confidence bound to rule out a 50% or lower true rate. (This result is not shown in the table; 65% favorable outcome rate of 48 subjects transplanted results in 31 successfully transplanted subjects). The proposed sample size is 48 subjects. Each participating center is expected to enroll at least 6 subjects.

Table 3 displays the probability that the study would conclude that the true favorable outcome rate is at least 50% for several selected values of the true favorable outcome rate. The tabulated probabilities are the power that a one-sided binomial test of the null hypothesis H_0: $p<0.5$ versus the alternative hypothesis H_a: $p \geq 0.5$ would conclude that the true favorable outcome rate is at least 50% given that the selected value is the true underlying favorable outcome rate. These values are obtained by calculating the binomial probability that at least 31 out of 48 total enrolled subjects would achieve the favorable outcome given the true outcome rate.

Table 3- The power to rule out a favorable outcome rate that is less than or equal to 50% for the given true favorable outcome rate.

True Favorable Outcome Rate	Power to Rule Out Favorable Outcome Rate
10%	0.0000
20%	0.0000
30%	0.0000
40%	0.0005
50%	0.0297
60%	0.3111
70%	0.8359
80%	0.9962
90%	1.0000

This table shows that if the true favorable outcome rate is 70%, the power of concluding that the rate is over 50% is 0.8359 for a 5% level test.

3. General Analysis Definitions

3.1 *Study Period and Visit Window Definition*

3.1.1 Study Period

The trial consists of three periods: (1) the pretransplant period, which includes screening, enrollment, and wait list time; (2) the period that includes the islet transplant procedure(s); and (3) follow-up visits through 24 months following the final transplant. The study period of this trial is a 24-month follow-up after the final islet transplant. Subjects may undergo up to 3 transplants in the course of this study; the final transplant can occur not later than 8 months following the first transplant.

In the first period, individuals who meet the general inclusion criteria will be approached regarding participation in the study. After informed consent has been obtained, they will be formally enrolled into the study. Eligibility will be confirmed based on the results of the screening visit procedure detailed in Appendix 1 of the CIT-07 protocol. Enrolled subjects who meet the eligibility screening for the studies will be put on the waiting list for an islet transplant.

Once a compatible islet preparation becomes available, a subject's eligibility will be re-confirmed and eligible subjects will begin immunosuppression therapy on Day -2 (Day 0 is defined as the day of transplant).

During the post-transplant follow-up period, subjects may receive up to two additional transplants. After receiving his/her initial islet transplant, if a subject does not meet the criteria for insulin independence described in Section 4.1.2 of the CIT-07 protocol, but has either a basal or stimulated C-peptide level $\geq$ 0.3 ng/mL (0.1 nmol/L), s/he will be considered for a second islet transplant. A second islet transplant will be considered at 75 $\pm$ 5 days after the first islet transplant and when all required metabolic assessments are complete. If, after the second islet transplant, both basal and stimulated C-peptide levels remain <0.3 ng/mL (0.1 nmol/L), the recipient will be considered to have failed the endpoint, and immunosuppression will be managed as described in the protocol. A third islet transplant will be considered only if all the criteria described in Section 7.4 of the CIT-07 protocol are met. Islet transplant recipients who have completed 12 months of follow up after their first infusion will no longer be eligible for additional islet transplants under the CIT-07 protocol.

3.1.2 Visit Windows

The number of visits that occur before the first islet transplant will be determined by time on the waiting list and cannot be determined in advance. Screening tests and baseline measurements that are obtained during this period must be obtained within specified windows relative to the day of randomization, as described in Appendix 1 of the CIT-07 protocol (section 7). Following a transplant, up to 13 visits may be scheduled. Table 4 describes all visits, their scheduled times relative to the islet transplant, and the allowable visit windows.

Table 4 - Study assessment time points and visit windows

Time Points (days relative to transplant)	Visit Number	Visit Window (days relative to transplant)	Equivalent (weeks or months)
Screening	01	N/A	N/A
Waitlist/ Baseline	02	N/A	N/A
0	03	N/A	N/A
3	04	N/A	N/A
7	05	±3	W1
14	06	±3	W2
21	07	±3	W3
28	08	±3	W4
56	09	±7	M2
75	10	±5	M2.5
120	11	±7	M4
150	12	±7	M5
180	13	±7	M6
270	14	±14	M9
365	15	±14	M12

This follow-up schedule will restart with visit 03 on the day that a subsequent islet transplant is performed. There is also a visit scheduled 365 days after the initial transplant. The detailed activities for each scheduled follow-up visit are described in Appendix 1 of the CIT-07 protocol.

3.2 *Study Population*

The study population consists of individuals with T1D who meet the eligibility criteria for the trial described in Section 3 of the CIT-07 protocol. This section of the SAP describes three study populations. All efficacy analyses will be done on the ITT population. Parallel analyses will be done on the per-protocol population. Safety analyses will focus on the safety population.

3.2.1 Intent-to-Treat Population

All efficacy and safety analyses will be based on a modified ITT principle: any subject in whom protocol-directed therapy (*e.g.*, immunosuppression) is initiated will be included in the ITT population. Subjects who are randomized but for whom a protocol directed therapy is not initiated will be listed in the final study report but will not be included in the ITT population.

A subject who is randomized but never receives protocol-directed therapy will not be included in the analysis.

3.2.2 Per-Protocol Population

A per-protocol analysis will include all subjects who are randomized to CIT-07 and in whom the islet transplant procedure is initiated. The procedure will be considered initiated when the operator (*e.g.*, surgeon or interventional radiologist) has started the process of obtaining access to the portal vein (*i.e.*, entered the body with a needle or scalpel).

3.2.3 Safety Population

The safety population consists of any subject in whom protocol-directed therapy (*e.g.*, immunosuppression) is initiated. Subjects in this population might not receive an islet transplant.

3.3 *Treatment Assignment and Treatment Groups*

3.3.1 Treatment Assignment

Enrolled subjects who meet the eligibility criteria will be placed on a waiting list for a transplant. Once a compatible pancreas becomes available, the subject will be reevaluated to ensure that s/he satisfies all inclusion/exclusion criteria and therefore is still eligible for CIT-07 and the site-specific Phase 2 study. Eligible subjects will be randomized into either the CIT-07 or the site-specific phase 2 study. Note that this randomization is between protocols and not to treatment arms within a protocol. This randomization is being performed to avoid bias in the assignment of subjects to protocols. The treating center will be blinded to protocol assignments until the subject is ready for transplantation.

3.3.2 Treatment Groups

The treatment in CIT-07 is the islet transplant and its associated immunosuppression. This is a single arm, open label trial and everyone assigned to this protocol receives the same study treatment.

3.3.3 Center Pooling Method

The primary analysis assumes no between-center variability with regard to the favorable outcome resulting from the treatment. The data from all centers will be pooled without any adjustment for centers.

After the completion of CIT-07, it may be of interest to evaluate each center's individual favorable outcome rate. For this secondary analysis, a Bayesian random-effects model, allowing for the possibility of variability between centers, will be implemented. A detailed analysis plan for the Bayesian random-effects model with implementation code is described in Appendix 8.

3.4 Subject Disposition

The number of subjects enrolled and treated will be summarized and reported in the following categories:

a. The numbers of subjects who are screened - total and grouped by center.

b. The numbers of subjects who are excluded from study participation - total and grouped by center. The numbers will also be tabulated by the reason for exclusion.

c. The numbers of subjects who are enrolled (sign informed consent for screening) - total and grouped by center.

d. The numbers of subjects who are eligible for the transplant both after the enrollment and before the transplant, grouped by center.

e. The numbers of subjects who are lost to follow-up, grouped by center and reason. (The rules for premature termination of study treatment are fully described in Section 5.7.2 of the CIT-07 protocol).

A list of all enrolled subjects (grouped by center) who are prematurely terminated from the study (withdraw consent or are lost to follow-up) will be provided. The list will give subject identification, the specific reason for termination, immunosuppression regimen and the duration of treatment before the termination.

3.5 *Protocol Deviations*

Major protocol deviations will be summarized by center and for the total study and grouped into the following categories:

1. Impacts the inclusion and/or exclusion criteria (PD1)
2. Involves consent violations (PD2)
3. Alters protocol-specified study therapy (PD3)
4. Impacts the ability of the Sponsor to evaluate the endpoints of the study (PD4)
5. Involves administration of prohibited medications (PD5)

The template of summary tables for the protocol deviations is provided in Appendix 6. Individual subjects with these protocol deviations will be listed with specifics on the deviation.

3.6 *Demographic and Baseline Characteristics*

3.6.1 Baseline Data

Baseline data collected for the ITT analysis consists of demographic information and medical/physical assessments during the waiting period for islet transplant. These data will be grouped into the following categories:

Demographic variables

- Age
- Sex
- Race (White, Black, Hispanic, Asian and Other)

Diabetes Control

- Insulin requirement (units/day)
- HbA1c

- Fasting and post prandial plasma glucose
- Fasting and post prandial C-peptide
- 90-minute C-peptide following consumption of Boost® or equivalent
- 90-minute glucose following consumption of Boost® or equivalent
- Number of severe hypoglycemic events in the last year
- MAGE score
- LI
- Clarke Score
- HYPO score
- β-score
- C-peptide/(glucose·creatinine) ratio

Body Habitus and Quality of Life

- Body Weight
- Height
- QOL Measures (defined in SAP section 3.6.5)

3.6.2 Statistical Analysis of Baseline Data

The number of subjects who do not meet the eligibility criteria will be reported, grouped by center. The numbers will be further broken down by the reasons for exclusion. These numbers and the corresponding rates will be reported in the summary table provided in SAP Appendix 7.

Descriptive statistics of baseline data, grouped by center, will be presented in a summary table. Continuous data will be summarized by mean, standard deviation, median, minimum, and maximum. Categorical data will be presented as numbers and percentages. The template of this table is described in SAP Appendix 1. Since the number of subjects randomized in each center is small, there will be no significance tests for the differences of the baseline data among the seven centers.

3.6.2.1 Analysis of demographic variables

3.6.2.1.1 *Age*

The distribution of age at baseline will be displayed as mean, standard deviation, median minimum and maximum.

3.6.2.1.2 *Sex and race*

Sex and race will be displayed as counts and percents.

3.6.3 Analysis of Diabetes Control Variables

Diabetic control variables include insulin requirement, fasting and post-prandial plasma glucose and C-peptide, glucose and C-peptide 90 minutes post Boost® (or equivalent) glucose challenge, number of hypoglycemic events in the last year, MAGE score, LI,

Clarke score, HYPO score, β-score, and C-peptide/(glucose·creatinine) ratio (the ratio of C-peptide to the product of glucose and creatinine).

The distribution of number of hypoglycemic events in the last year will be presented in tabular form.

The remaining diabetes control variables are all continuous scale variables and will be displayed as mean, standard deviation, median, minimum and maximum. The Shapiro-Wilk test will be used to test the normality of this continuous variable. If the hypothesis of normally distributed data is not rejected at the 0.05 significance level, the usual normal 95% confidence interval $\bar{x} \pm t_{0.025}^{n-1} \times \frac{s_x}{\sqrt{n}}$ will be constructed for the true mean μ. If the data are not normally distributed, the transformations $y = f(x) = \log x, \sqrt{x}$ and $1/x$ of these variables will be examined sequentially until we fail to reject the hypothesis of normality using the Shapiro-Wilk test at the 0.05 significance level. If a transformation is identified such that normality holds then the sample mean $\bar{y}$ and sample standard deviation s_y will be calculated for the transformed variable. The 95% confidence interval will be constructed as

$$\left[f^{-1}\left(\bar{y} - t_{0.025}^{(n-2)} \frac{s_y}{\sqrt{n}} \right), f^{-1}\left(\bar{y} + t_{0.025}^{(n-2)} \frac{s_y}{\sqrt{n}} \right) \right].$$

If these transformations fail to achieve normality, the bootstrap method will be adopted. One thousand bootstrap samples from the original data set with replacement are obtained. For each bootstrap sample, the sample mean is calculated. The 2.5 and 97.5 percentiles, $p_{0.025}$ and $p_{0.975}$, will be identified from the 1000 bootstrap sample means. The interval $[p_{0.025}, p_{0.975}]$ will reported as the bootstrap 95% confidence interval for the true mean.

3.6.4 Analysis of Body Habitus Variables

Body habitus variables include height, weight, and body mass index (BMI). BMI is computed as the ratio of weight to the square of height (kg/m^2). All three variables are continuous scale and will be analyzed in the same manner as described for continuous variables in the previous section.

3.6.5 Analysis of Quality of Life Variables

In this study, five QOL measures are considered. Two scales are obtained from the SF36 questionnaire (version 2). These scales are the Summary Physical Component score (SPC) and the Summary Mental Component score (SMC). Additional QOL assessments include the Overall Health Status (OHS) measure based on the European Quality of Life (EQ-5D) instrument, the Diabetes Distress Scale (DDS), and the Hypoglycemia Fear Scale (HFS) scales.

The SF36 scales will be standardized to the US population. The EQ-5D OHS questionnaire provides two sub-scales: the Behavior subscale that is based on 10 questions and ranges from 0 to 40 and the Worry subscale that is based on 13 questions. The scale is computed as the sum of the responses to the 13 questions and ranges from 0 to 52. The DDS is based on 17 questions rated on a 6 point Likert scales and ranges from

17 to 102. The HFS has two subscales: the Behavior scale which ranges from 0 to 40 and the Worry scale that ranges from 0 to 52.

All QOL scales are continuous and will be analyzed in the manner described in SAP section 3.6.3 for the continuous scale diabetes control variables. Each scale will be tabulated as mean, standard deviation, median, minimum and maximum. 95% confidence intervals will be computed in the manner described in SAP section 3.6.3.

3.7 *Efficacy Analyses*

3.7.1 Statistical Analysis of Primary Endpoint

HbA1c is the standard measure of glucose control and is used in all major studies as an endpoint for glycemic control. It has been valuable as a risk predictor of diabetes complications. However, since HbA1c is an integrated average, it does not provide information about the range of glucose values a subject experiences. This limitation is a rationale for also including hypoglycemic event occurrence as part of the primary endpoint and various glycemic excursion measures as secondary endpoints.

The primary endpoint for this study is the proportion of subjects with an HbA1c <7.0% at Day 365 AND free of severe hypoglycemic events from Day 28 to Day 365, inclusive, following the first islet transplant, with the day of transplant designated Day 0.

The primary analysis is designed to estimate the true rate of favorable outcomes at one year in subjects in the ITT population pooled over all centers. The proportion of favorable outcomes will be used as the point estimate. An exact one-sided 95% confidence interval will be constructed assuming an underlying binomial distribution for the target population as follows:

If r out of 48 total enrolled subjects achieve the favorable outcome, the exact one-sided 95% confidence interval is given by $p \geq p_L$ where p_L is the solution of the equation

$$0.05 = \sum_{x \geq r} \binom{48}{x} p_L^x (1 - p_L)^{48-x}.$$

$0.05 =$

The rates of favorable outcome and exact one-sided 95% confidence intervals will be computed for each contributing center.

This analysis will be conducted for the ITT population. The primary endpoint should be available for all treated subjects. An exception will be if a death occurs, if the subject withdraws consent to be followed, or if immunosuppression is begun but the subject never receives a transplant. In these cases the endpoint will be classified as failure to achieve a favorable outcome. Should the endpoint not be evaluated for a particular individual for other reasons, a failure will be imputed unless an evaluation is done at a time longer than one year after transplant, in which case, that later value will be imputed. All imputations will be reported with the primary analysis. The rates and the exact one-sided 95% confidence intervals for complete data and imputed data will be compared to ascertain the sensitivity of the imputation.

3.7.2 Statistical Analyses of Secondary Endpoints

Except for the primary analyses there are no explicit or implied hypotheses in the protocol. Changes in the secondary outcomes are of interest as they will relate to efficacy as measured by the primary outcome variable. All analyses are descriptive and are intended to document the changes in these important variables. Secondary outcomes will be used to support the decision for efficacy of islet transplantation in this population but are not intended to be used explicitly for making a decision for the efficacy of islet transplantation in this population.

3.7.2.1 Subgroup analysis

CIT-07 is a multi-center single arm clinical trial to estimate the probability of achieving the endpoint at one year after enrollment (called "favorable outcome"). The primary analysis is described in SAP Section 3.7.1 and assumes no between-center variability: the response is a single binomial response of the total number of favorable outcomes among the total number of subjects enrolled. This assumption in the primary analysis of no between-center variability implies that the underlying success probability is the same at each center. Data from CIT centers are consistent with this assumption of no between center variability. In addition, the specification of a common protocol for manufacturing the islets and a common protocol for transplantation are also consistent with there being no between-center variability.

However, in order to examine the assumption of no between-center variability and to assess the rate of favorable outcome at each center, a Bayesian random effects model will be implemented. The rate of favorable outcome at each of the centers preparing islet cells for transplant will be estimated as a planned subgroup analysis using a Bayesian random-effect model under the assumption that the favorable outcome rates are *a priori* exchangeable. That is before any subjects are enrolled the centers are assumed to be similar (exchangeable) but, after data are analyzed, may be different. A one-sided 90% probability interval will be constructed to estimate a 90% posterior probability lower bound for efficacy at each center. The following section describes the methods to be used and provides supporting power calculations. A complete description, including the programs, is given in SAP Appendix 8.

This subgroup analysis will have no bearing upon the efficacy and safety analyses conducted for the study as a whole. Therefore, Bayesian analysis will not be considered in the adjustments for multiplicity that are described later for the "Key Secondary Endpoints".

The hierarchical nature of the model introduces a correlation structure between centers which allows the estimates for each manufacturing center to incorporate data from other manufacturing centers. Smaller standard errors will therefore be provided than through using only data from each center in isolation from data from other centers. The estimates are adjusted towards the average success rate in all centers (also known as shrinkage estimators). The adjustment to the overall success rate depends on the variability observed in the actual data realized in the study and the sample sizes in each center.

If the data from the centers are similar to each other, then this is consistent with there not being very much between-center variability and there will be more shrinkage to the

overall analysis. If any of the centers are different from the others and have a different rate of favorable outcome the analysis will adapt, and there will be less shrinkage to the average rate of favorable outcome.

The trial data will be analyzed using the WinBUGS program [15]. Slightly different analyses result depending on exactly what prior assumptions are made. The distribution on the between-center variability is important to specify carefully, because default improper prior distributions are not appropriate and lead to unstable estimation. A prior distribution on the between center variability therefore is assumed. The particular prior distribution chosen, a gamma distribution on the inverse of the variance, is not very informative, but contains enough information to make the analysis robust and stable. The prospective data from CIT-07 will therefore stand alone in all but the assumptions on the variability between centers.

The hypothetical data sets in Appendix 8 were analyzed using WinBUGS. Unless stated otherwise, the simulations in this section and Appendix 8 were generated using the R program [16] and OpenBUGS [17], with the BRugs interface to call OpenBUGS from R. In a few cases (involving extreme cases where individual centers may have results of 100% or 0% favorable outcomes), R2WinBUGS was used to avoid some complications where the sampler failed to satisfy convergence diagnostics, in which case those cases were not included in the calculations (less than 2% of the simulations).

3.7.2.1.1 *The model and notation*

Suppose there are k centers and each center recruits n_i subjects, $i=1,..,k$: it is anticipated that $k=6$ and $(n_1, n_2, n_3, n_4, n_5, n_6) = (12, 12, 6, 6, 6, 6)$. Let y_i denote the observed number of favorable outcomes at center i out of n_i subjects, and let p_i denote the true underlying success probability at the *ith* center. The random-effects framework postulates that the observed number of successes for each center, y_i, conditional on p_i, is a draw from a binomial distribution, independently at each center:

$y_i|p_i$ is Binomial(n_i, p_i)

for $i=1,...,k$. Further denote the log odds of success at each center by θ_i

$\theta_i = log[p_i/(1-p_i)]$

for $i=1,...,k$. Furthermore, assume that there may be between-center variability, but all centers are exchangeable: let θ_i given μ and σ^2 be like a sample from a normal population:

$\theta_i|\ \mu, \sigma^2$ is N(μ, σ^2).

The values μ and σ represent the population log odds and population between-center standard deviation of the log-odds respectively.

Let $\pi=e^{\mu}/(1+e^{\mu})$, the inverse log odds transformation, then π is the population probability of favorable outcome, the underlying predicted success probability for the population of centers.

A Bayesian approach, using Monte Carlo Markov Chain (MCMC) methods, will be implemented to estimate parameters and hyperparameters. The parameters of direct

interest are: the overall probability of favorable outcome π, and each of the probabilities of favorable outcome at individual centers p_i, $i=1,\ldots,k$.

3.7.2.1.2 *The prior distribution*

The analysis of the CIT-07 data uses a very vague prior distribution with the distribution of μ assumed uniform and the distribution of σ^{-2}, denoted as τ, assumed to have an independent and proper gamma distribution with a large standard deviation. Specifically, τ has a density proportional to $\tau^{\alpha-1} e^{-\beta\tau}$ *for* $\tau>0$ [18, p. 39]: values of $\alpha=2$ and $\beta=1.5$ are specified for analysis. These values lead to stable estimation and are consistent with having very little prior information on the between center variability [19]. The location of the prior mean of the precision is also approximately consistent with historical data and the variance has been inflated to represent uncertainty.

The gamma prior distribution for τ has a mean of α/β and variance of α/β^2. The distribution of the θ_i's given μ is a scaled t-distribution [18, p. 42]. If, for example μ corresponds to a mean for all centers of $\pi = 0.7$, then the conditional prior distribution of any of the p_i (that is the distribution of $p_i|\mu$ is such that it lies between 0.17 and 0.96 with probability 0.95, see SAP Appendix 8.1).

An additional setting of $\alpha=2$ and $\beta=0.75$ is examined in detail in Appendix 8.1. This gives a smaller prior mean value for the between center variability, which typically leads to more shrinkage towards the overall mean.

3.7.2.1.3 *Estimation of rate at individual centers*

Twelve subjects at one center is considered sufficient to demonstrate comparable efficacy at that center. This is based on the power calculations given in SAP Appendix 8.1.

Two methods for the estimation of the rate at individual centers are used: 1) estimation based on the complete set of endpoint data from CIT-07 centers using data from CIT protocols 02, 03, 04, 05, 06, and 07 and 2) estimation based on the endpoint data from CIT-07 centers using data from CIT-07 alone.

3.7.2.1.3.1 **Estimation of the rate at individual centers based on the complete set of endpoint data from CIT-07 centers using data from CIT protocols 02, 03, 04, 05, 06, and 07**

Those CIT-02, -03, -04, -05, -06, and -07 subjects enrolled at a CIT-07 participating center are assessed for meeting the CIT-07 primary endpoint definition. If a center enrolls at least 12 subjects from the combination of the CIT-02, -03, -04, -05, -06, and -07 protocols, the criteria for evaluations of that center require that the resulting analysis satisfy both (a) and (b) below:

a) The overall primary result of the combined protocols is positive: that is, the primary analysis, which constructs a one-sided 95% confidence interval for the overall probability of favorable outcome assuming no between center variability, rules out a favorable outcome rate of 0.50 or less.

b) The favorable outcome rate at that center, under the Bayesian model, is estimated to be at least 0.45 with posterior probability 0.90. That is, the lower 90% probability bound is at least 0.45.

3.7.2.1.3.2 Estimation of the rate at individual centers based on the endpoint data from CIT-07 centers using data from CIT-07 alone

The subgroup analysis described in SAP section 3.7.2.1.1 will also be fit to a dataset comprised of the CIT-07 subjects alone. If a center enrolls at least 12 subjects in CIT-07, the criteria for evaluations of that center require that the resulting analysis satisfy both (a) and (b) below:

a) The overall primary result of CIT-07 is positive: that is, the primary analysis, which constructs a one-sided 95% confidence interval for the overall probability of favorable outcome assuming no between center variability, rules out a favorable outcome rate of 0.50 or less.

b) The favorable outcome rate at that center, under the Bayesian model, is estimated to be at least 0.45 with posterior probability 0.90. That is, the lower 90% probability bound is at least 0.45.

3.7.2.1.4 Power calculations (α=2, β=1.5)

For any fixed value of the p_is, binomial data can be simulated and the results analyzed and examined as to whether the criteria above are met. This will give an estimate of "power" where power is defined as the probability of meeting the criteria under various hypothesized parameter values.

For this section, we assume the sample sizes are two centers each with 12 subjects and 4 centers each with 6 subjects, for a total of n=48 subjects in CIT-07.

Preliminary data from CIT centers indicates a rate of favorable outcome between 70% and 90%.

For illustration therefore, suppose that $\pi = 0.70$ and there is no between center variability and so $p_1=p_2=p_3=p_4=p_5=p_6=0.70$. A total of 10,000 data sets were simulated and the proportion where criterion 1 was satisfied was calculated for an estimate of global power, as was the proportion where an individual center satisfied both criterion (a) and criterion (b) for an estimate of the individual center power. The individual power depends on the sample size in each group. The value for a centers with 12 subjects is 0.77 and is denoted as the individual power for a sample of size 12. A second simulation of 10,000 data sets was generated for $\pi = 0.70$ and gave the same result, and therefore 10,000 data sets were generated for other values of π. Results for the global power and the individual center power are given in Table 5.

Table 5 - Power and Size Estimation for α=2 and β=1.5

Probability of Favorable Outcome at Each Center	Global Power	Individual Power (n=12)
0.5	0.03	0.025
0.7	0.84	0.77
0.8	1.00	0.98
0.9	1.00	1.00

The entries in this table can be interpreted as follows. If the true rate is 0.50 or less, there

is an appropriately low probability (0.025 or less) of a center meeting the criteria (corresponding to the concept of size or a type I error). If there is a high rate of favorable outcome, there is a high probability of meeting the criteria (corresponding to power): for a probability of 0.70, there is an individual center power of 77% and for 0.80, an individual center power of 0.98.

The above calculations were done calling WinBUGS from R and sample code is available in Appendix 8.2. Each set of 10,000 simulations takes approximately 12 hours of computer time. The corresponding probabilities that the analysis indicates that the centers with 6 subjects meet criteria 1 and the lower 90% posterior probability bound is greater than 0.45 is is 0.02 and 0.66 for true probabilities of 0.50 and 0.70 respectively. These are less than the individual power for the centers enrolling 12 subjects.

3.7.2.1.5 *Additional power calculations*

For completeness additional simulations were run assuming different true rates of favorable outcomes at each center. In all cases, $n_1=n_2=12$ and $n_3=n_4=n_5=n_6=6$ was used as the sample size: two centers enrolling 12 subjects and 4 centers enrolling 6 subjects each.

Centers enrolling 6 subjects in CIT-07 and 6 in a different CIT protocol will have an individual power slightly higher than a center enrolling 12 subjects in CIT-07 with the same true probability of favorable outcome, but no more than the Global Power (see SAP section 3.7.2.1.3.1 for the analysis in these cases).

Table 6 - Additional Power Calculations for α=2 and β=1.5 (The final 4 entries in this table were calculated using R2WinBUGS)

N=(12,12,6,6,6,6)	Global Power	Individual Power (n=12)
P=(0.7,0.7,0.7,0.7,0.5,0.5)	0.59	0.56
P=(0.7,0.7,0.7,0.7,0.7,0.5)	0.72	0.68
P=(0.7,0.7,0.7,0.7,0.7,0.3)	0.60	0.56
P=(0.5,0.7,0.7,0.7,0.7,0.7)	0.60	0.31 (center 1) 0.57 (center 2)
P=(0.3,0.7,0.7,0.7,0.7,0.7)	0.29	0.03(center 1) 0.29(center 2)
P=(0.9,0.9,0.9,0.9,0.5,0.5)	1.00	1.00
P=(0.9,0.9,0.9,0.9,0.9,0.5)	1.00	1.00
P=(0.5,0.9,0.9,0.9,0.9,0.9)	1.00	0.40 (center 1) 1.00 (center 2)
P=(0.3,0.9,0.9,0.9,0.9,0.9)	0.99	0.04 (center 1) 0.99 (center 2)

The detailed description of this subgroup analysis plan is provided in SAP Appendix 8.1.

3.7.3 Analysis of Key Secondary Endpoints

Because this is a single intervention study, the tests will be one sided tests for whether

the true rates are greater than the endpoint's' predetermined "minimum rate for efficacy". These minimum rates were determined by the investigators to be large enough to have credibility for the islet transplant community. The minimum rate for efficacy is provided in Table 7 for each of the key secondary outcomes.

Table 7: Key Secondary Outcomes and Minimum Rates for Efficacy

	Key Secondary Outcome	Minimum* Rate for Efficacy
1	The proportion of subjects with an HbA1c <7.0% AND free of severe hypoglycemic events from Day 28 to Day 730, inclusive, after the first islet transplant.	50%
2	The proportion of subjects with HbA1c ≤ 6.5% at one year after the first islet transplant AND free of severe hypoglycemic events from Day 28 to Day 365 after the first transplant.	50%
3	The proportion of subjects with HbA1c ≤ 6.5% at two years after the first islet transplant AND free of severe hypoglycemic events from Day 28 to Day 730 after the first transplant.	40%
4	The proportion of subjects free of severe hypoglycemic events from Day 28 to Day 365 after the first islet transplant.	50%
5	The proportion of subjects free of severe hypoglycemic events from Day 28 to Day 730 after the first islet transplant.	40%
6	The proportion of subjects with HbA1c <7.0% at one year after the first islet transplant.	50%
7	The proportion of subjects with HbA1c <7.0% at two years after the first islet transplant.	40%
8	The proportion of subjects with HbA1c ≤6.5% at one year after the first islet transplant.	50%
9	The proportion of subjects with HbA1c ≤6.5% at two years after the first islet transplant.	40%
10	The proportion of insulin-independent subjects at one year after the first islet transplant.	20%
11	The proportion of insulin-independent subjects at two years after the first islet transplant.	10%

As with the primary endpoint, the key secondary endpoints should be available for all transplanted subjects and the analysis will be conducted for the ITT population. If an endpoint is not available for a randomized subject then it will be imputed using the same rules that were used for the primary endpoint.

The observed rate for each key secondary outcome will be used as the point estimate. The analysis will also compute an exact binomial one-sided test for the null hypothesis that the true rate of the outcome is less than or equal to the predetermined minimum rate for efficacy against the alternative that the true rate exceeds the minimum rate for efficacy.

We will use the Benjamini and Hochberg [20] method to account for the multiplicity of the key secondary tests. This method controls the false discovery rate (FDR) rather than the more familiar family-wise error rate (FWER). It provides a powerful approach for identifying those positive tests that are not likely to be true while controlling a reasonable measure of the expected number of false positive tests (the FDR).

The multiple testing procedure considers testing m hypotheses $H_1, H_2, \ldots, H_m$. Each test yields the corresponding p-values $P_1, P_2, \ldots, P_m$. Let $P_{(1)} \leq P_{(2)}, \ldots, \leq P_{(m)}$ be the ordered p-values, and denote $H_{(i)}$ denote the null hypothesis corresponding $P_{(i)}$. The Bonferroni-type multiple testing procedure is defined by the following:

Let q^* the maximum false discovery rate and let k be the largest i for which $P_{(i)} \leq \frac{i}{m} q^*$; then reject all $H_{(i)}$ $i=1,2,\ldots,k$.

For this study there are 11 key secondary endpoints so m=11 and we will fix q^* at 0.1.

For each key secondary outcome, the observed rate will be used as the point estimate. The rate and a 95% exact one-sided confidence interval will be reported along with the p-value from an exact one-sided test of the corresponding null hypothesis.

3.7.4 Additional Efficacy Endpoints

The 11 additional efficacy endpoints are just the key secondary endpoints measured at one and two years following the final islet transplant (the key secondary endpoints are measured at one and two years after the initial transplant). They will be analyzed in exactly the same way as the key secondary endpoints but no adjustment will be made for multiple comparisons.

3.7.5 Analysis at 75 ± 5 Days Following the Initial and Final Infusion(s)

The analysis will be conducted for the ITT population at 75±5 days after the initial islet infusion. Separate analyses will be conducted for each variable observed at 75±5 days after the initial islet infusion and at 75±5 days after the final islet infusion. These times can differ by as much as 8 months for individual islet recipients.

3.7.5.1 Analysis templates

Two analytic templates are described as follows:

1) **Binary outcome variables:** For a binary variable, we will calculate the sample proportion based on the pooled sample as the point estimate of the true rate. The exact binomial 95% confidence interval is constructed for the true rate.
2) **Continuous scale outcome variables:** For a continuous variable, the sample mean $\bar{x}$ and sample standard deviation s_x are calculated based on the pooled sample. The sample mean $\bar{x}$ is used as a point estimate of the true mean μ. The Shapiro-Wilk test will be used to test the normality of this continuous variable. If the hypothesis of normally distributed data is not rejected at the 0.05 significance level, the usual

normal 95% confidence interval $\bar{x} \pm t_{0.025}^{n-1} \times \frac{s_x}{\sqrt{n}}$ is constructed for the true mean μ. If the data are not normally distributed, the transformations $y = f(x) = \log x, \sqrt{x}$ and $1/x$ of this variable will be examined sequentially until we fail to reject the hypothesis of normality using the Shapiro-Wilk test at the 0.05 significance level. If a transformation *f* is identified such that normality holds (*i.e.*, *normality acceptable for the transformed data y=f(x)),* then the sample mean $\bar{y}$ and sample standard deviation s_y are calculated for the transformed variable. The 95% confidence interval will be constructed as

$$\left[f^{-1}\left(\bar{y} - t_{0.025}^{(n-2)} \frac{s_y}{\sqrt{n}} \right), f^{-1}\left(\bar{y} + t_{0.025}^{(n-2)} \frac{s_y}{\sqrt{n}} \right) \right]$$

If these transformations fail to achieve normality, the bootstrap will be adopted. 1000 bootstrap samples sampled from the original data set with replacement are obtained. For each bootstrap sample, the sample mean is calculated and then the 2.5 and 97.5 percentiles, $p_{0.025}$ and $p_{0.975}$, will be identified from the 1000 bootstrap sample means. The interval $[p_{0.025}, p_{0.975}]$ will reported as the bootstrap 95% confidence interval for the true mean.

Analysis template tables for the following efficacy endpoints at 75±5 days following the initial and final infusion are provided in SAP Appendix 9.

3.7.5.2 The percent reduction in insulin requirement

Subjects will record their total daily insulin dose on self-monitoring diaries. The insulin requirement will be evaluated at 75±5 days following the initial infusion (and final infusion if applicable). The reduction percentages from the baseline insulin requirement at both times for a subject, if applicable, are calculated accordingly. The analysis of the reduction percentage adopts the method for a continuous variable.

3.7.5.3 HbA1c

HbA1c is measured at 75±5 days following the initial and final infusion (if applicable). The analysis of this measure adopts the method for a continuous scale variable. Data will be displayed as mean, standard deviation, minimum and maximum. In addition, 95% confidence intervals will be computed using an appropriate technique. In addition, change from baseline will be computed. Descriptive statistics and confidence intervals will be computed using the methods used for the original variable.

3.7.5.4 Mean amplitude of glycemic excursions (MAGE)

The MAGE requires capillary glucose readings over two consecutive days (a minimum of four readings a day), and is defined as the arithmetic mean of blood sugar increases (or decreases) when both increases and decreases (or vice-verse) at subsequent points in time are greater than 1 standard deviation of the blood sugar for the same two day period [2]. If the MAGE is ≥11.1 mmol/L, the subject is considered to have labile diabetes. The

MAGE will be measured at 75±5 days following the initial and final infusion (if applicable).

The first analysis for this variable directly uses MAGE and adopts the method for a continuous scale variable. Subjects will also be categorized into two groups depending on whether or not the MAGE ≥ 11.1 mmol/L. The analysis of the rate of labile diabetes will be displayed using the methods described for a binary outcome variable.

3.7.5.5 Glycemic lability index

LI is a measure of lability which is based on the change in glucose over time. The LI requires 4 or more daily capillary blood glucose (BG) measurements over a 4 week period. For each week, the sum of the squared differences in consecutive glucose readings is divided by the hours apart the readings are determined. Only differences in time that were greater than or equal to 1 hour and less than or equal to 12 hours are used in calculations. This sum is calculated for each of the four weeks and the LI is the mean of the four weekly values [3], that is:

$$\text{LI}=\frac{1}{4}\times\sum_{i=1}^{4}\sum_{j=1}^{N_i}\frac{(\Delta Gluc_{i,j})^2}{\Delta h_{i,j}}$$

where $\Delta Gluc_{i,j}$ is the j^{th} eligible difference of glucose readings in the i^{th} week, $\Delta h_{i,j}$ is the time interval in hours for the j^{th} eligible difference of glucose readings in the i^{th} week. N_i is the total number of the eligible differences of glucose readings in the i^{th} week which may vary from week to week.

Most subjects have scores under 300 mmol/L^2/h·wk^{-1} with a median of 223 (25 – 75$^{\text{th}}$ percentiles 130 – 329 mmol/ L^2/h·wk^{-1}). An LI $\geq$433 mmol/ L^2/h·wk^{-1} (90$^{\text{th}}$ percentile) indicates serious problems with glycemic lability. The LI will be measured at 75 ± 5 days following the initial and final infusion (if applicable). The first analysis for this variable directly uses LI and adopts the method described for continuous scale variables. Change from baseline will also be computed and analyzed using the same methods.

Subjects will also be categorized into two groups depending on whether or not the LI $\geq$433 mmol/ L^2/h·wk^{-1}. The analysis of the rate of serious problems with glycemic lability will also be analyzed using the method for binary outcomes.

3.7.5.6 Ryan hypoglycemia severity (HYPO) score

The HYPO score involves subject recording of blood glucose readings and hypoglycemic events (BG <3.0 mmol/L [54 mg/dL]) over a 4-week period and recall of all severe hypoglycemic episodes in the previous 12 months. The HYPO score is a scalar quantity based on the severity of hypoglycemic events over a four week period. A hypoglycemic event occurs when a blood sugar reading is less than 54 mg/dL and a series of self-reported questionnaire items determine the severity. The HYPO score is the sum of points awarded to each hypoglycemic event, where a large HYPO score indicates more severity [3].

A HYPO score greater than or equal to the 90$^{\text{th}}$ percentile (1047) of values derived from an unselected group of type 1 diabetic subjects indicates severe problems with

hypoglycemia. The HYPO score will be measured at 75±5 days following the initial and final infusion (if applicable).

The analysis of the raw HYPO score adopts the method for continuous scale variable. Change from baseline will also be calculated and analyzed using the same method. Subjects with HYPO scores greater than or equal to 1047 will be classified as having severe problems with hypoglycemia, and the analysis of the true rate of this event adopts the method for binary outcome variables.

3.7.5.7 Basal (fasting) and 90-min glucose and C-peptide derived from the mixed-meal tolerance test

The partial graft function of islet transplantation is indicated by continued C-peptide production. Basal and 90-min glucose and C-peptide derived from the MMTT will be measured at 75±5 days following the initial and final infusion (if applicable). The glucose levels (fasting and 90 minute) are continuous variables and will be analyzed using the method for continuous scale variables. Change from basal to 90 minutes will be calculated and analyzed using the same methods. Change from baseline for basal, 90 minutes and the difference between 90 minutes and basal will also be analyzed using the same methods.

3.7.5.8 β-score

The β-score is an assessment of β-cell graft function after islet transplantation and is treated as a continuous variable and ranges from 0 (no graft function) to 8.

The β-score is generated from a composite scoring system based on fasting plasma glucose values (mmol/L), HbA1c(%), daily insulin consumption (units/kg) or oral hypoglycemic agents use, and stimulated C-peptide levels (nmol/L). For each of these measures, scores of 0, 1, and 2 are assigned to abnormal, intermediate, and normal values, respectively. The β-score is then calculated as the sum of the four scores [4].

The β-score will be measured at 75±5 days following the initial and final infusion (if applicable) and will be analyzed using the method for a continuous variable.

3.7.5.9 C-peptide/(glucose·creatinine) ratio

This measure accounts for both the dependence of C-peptide secretion on the ambient glucose concentration and the dependence of C-peptide clearance on kidney function [10, 11]. This ratio is a continuous variable and will be measured at 75±5 days following the initial and final infusion (if applicable). The analysis of this variable adopts the method for a continuous variable.

3.7.5.10 Acute insulin response to glucose (AIR_{glu}), insulin sensitivity, and disposition index derived from the insulin-modified frequently-sampled intravenous glucose tolerance (FSIGT) test

The insulin modified FSIGT test will be performed at 75±5 days following the initial and final infusion (if applicable). AIR_{glu}, SI and DI, derived from the test, provide a composite measure of β-cell function. AIR_{glu} is calculated as the incremental area-under-the-curve for insulin between 0 and 10 minutes post injection. SI, a measure of insulin-

dependent glucose disposal, is derived from Bergman's minimal model using MinMod Millenium software. The DI is calculated by AIR_{glu}*SI. The three variables are all continuous and analyzed using the method for a continuous variable.

3.7.5.11 Glucose variability and hypoglycemia duration derived from the continuous glucose monitoring system (CGMS)

CGMS involves the subcutaneous placement of a glucose sensor connected by tubing to a pager-sized monitoring device that stores glucose data obtained every 5 minutes over a 72-hour period. Data from a 72-hour period at 75±5 days following the initial and final infusion (if applicable) will be used to derive the glucose variability and hypoglycemia duration.

The glucose variability is the absolute value of measured glucose minus 5.5 mmol/L [100mg/dL] and is a continuous variable. This variable will be analyzed using the method for a continuous variable. The data from the 72-hour period are used to derive the number and duration of all hypoglycemic episodes (measured glucose <3.0 mmol/L [54 mg/dL]). Then the total duration of hypoglycemia can be calculated. The hypoglycemia duration is a continuous variable and is analyzed using the method for a continuous variable. The distribution of the number of hypoglycemic events will be tabled.

3.7.5.12 Quality of life measure

The analysis of QOL is deferred to SAP section 3.9.4.

3.7.6 Analysis at 365±14 Days Following the Initial and Final Infusion

The analysis will be conducted for the ITT population. The analysis template tables for the following efficacy endpoints at 365±14 days following the initial and final infusion (if applicable) are provided in Appendix 10.

3.7.6.1 The percent reduction in insulin requirement

The insulin requirement will be evaluated at 365±14 days following the initial and final infusion (if applicable). The reduction percentages from the baseline insulin requirement at both times for a subject, if applicable, are calculated accordingly. The analysis of the reduction percentage adopts the method for a continuous variable.

3.7.6.2 HbA1c

HbA1c is measured at 365±14 days following the initial and final infusion (if applicable). The analysis of this measure adopts the method for a continuous variable. Change from baseline will also be calculated and analyses will be conducted using the same method.

3.7.6.3 MAGE

The MAGE will be measured at 365±14 days following the initial and final infusion (if applicable). The first analysis for this variable directly uses MAGE and adopts the method for a continuous variable. Subjects will be categorized into two groups depending on whether the MAGE at 365±14 days post-transplant is >11.1 mmol/L or not.

The analysis of the rate of labile diabetes adopts the method for a binary outcome variable.

3.7.6.4 Lability Index

The LI will be measured at 365±14 days following the initial and final infusion (if applicable). The first analysis for this variable directly uses LI and adopts the method for a continuous variable. Subjects will be categorized into two groups depending on whether the LI $\geq$433 mmol/ $L^2/h \cdot wk^{-1}$ or not. The analysis of the rate of serious problems with glycemic lability adopts the method for a binary variable.

3.7.6.5 Clarke Score

The Clarke score is given by the investigator according to the Clarke survey questionnaire that gives a total score between 0 and 7 (most severe), where scores of 4 or more indicate reduced awareness of hypoglycemia and increased risk of severe hypoglycemic events. The Clarke score will be evaluated at baseline and at 365±14 days following the initial and final infusion (if applicable). The primary analysis of this variable will treat this variable as continuous and adopts the method for a continuous variable. Change from baseline will also be calculated and analyzed using the same method. For each subject, the indicator of high risk of severe hypoglycemic events will be created according to whether the Clarke score is greater than or equal to 4. The high risk rate of severe hypoglycemic events will be analyzed using the method for a binary outcome variable.

3.7.6.6 HYPO score

The HYPO score will be measured at 365±14 days following the initial and final infusion (if applicable). The analysis of the raw HYPO score adopts the method for a continuous variable. Change from baseline will also be calculated and analyzed using the same method. Subjects with HYPO scores greater than or equal to 1047 will be classified as having severe problems with hypoglycemia, and the analysis of the true rate of this event adopts the method for a binary outcome variable.

3.7.6.7 Basal (fasting) and 90-min glucose and C-peptide (MMTT)

Basal and 90-min glucose and C-peptide derived from MMTT will be measured at 365±14 days following the initial and final infusion (if applicable).

The partial graft function of islet transplantation is indicated by continued C-peptide production. Basal and 90-min glucose and C-peptide derived from MMTT will be measured at 365±14 days following the initial and final infusion (if applicable). The glucose levels (fasting and 90 minute) are a continuous variable and will be analyzed using the method for continuous scale variables. Change from basal to 90 minutes will be calculated and analyzed using the same method. Change from baseline for basal, 90 minutes and the difference between 90 minutes and basal will also be analyzed using the same method.

3.7.6.8 β-score

The β-score will be measured at 365±14 days following the initial and final infusion (if applicable) and is analyzed using the method for a continuous variable. Change from baseline will be calculated and analyzed using the same method.

3.7.6.9 C-peptide/(glucose·creatinine) ratio

This measure accounts for both the dependence of C-peptide secretion on the ambient glucose concentration and the dependence of C-peptide clearance on kidney function [10, 11]. This ratio is a continuous variable and will be measured at 365±14 days following the initial and final infusion (if applicable). The analysis of this variable adopts the method for a continuous variable.

3.7.6.10 Acute insulin response to glucose, insulin sensitivity, and disposition index derived from the insulin-modified frequently-sampled intravenous glucose tolerance test

The insulin modified FSIGT test will be performed at 365±14 days following the initial and final infusion (if applicable). AIR_{glu}, SI and DI, derived from the test, provide a composite measure of β-cell function. AIR_{glu} is calculated as the incremental area-under-the-curve for insulin between 0 and 10 minutes post injection. SI, a measure of insulin-dependent glucose disposal, is derived from Bergman's minimal model using MinMod Millenium software. The DI is calculated by AIR_{glu}*SI. The three variables are all continuous and analyzed using the method for a continuous variable.

3.7.6.11 QOL

See the detailed description in SAP section 3.9.4

3.7.6.12 The proportion of subjects receiving a second islet transplant

For each subject, whether the second islet transplant is implemented and will be observable at 365±14 days following the initial and final infusion (if applicable). The true rate of subjects who need a second islet transplant will be analyzed using the method for a binary outcome.

3.7.6.13 The proportion of subjects receiving a third islet transplant

For each subject, whether the third islet transplant is implemented will be observable at 365±14 days following the initial and final infusion. The true rate of subjects who need a third islet transplant will be analyzed using the method for a binary outcome.

3.7.6.14 Proportion of subjects with HbA1c<7.0% and free of severe hypoglycemic events at each center preparing islets

Rates and 95% exact confidence intervals will be reported for each individual center. Rates will be reported at one year and two years after the final transplant. See subgroup analysis (SAP section 3.7.2.1) for the description of a more in depth analysis that may be

used by the individual centers to support licensure of individual islet preparation laboratories.

3.7.7 Analysis at 730±14 Days Following the Initial and Final Infusion

Blood sugar readings are not required more than one year after the final islet transplant. Therefore, we will not be able to compute the Mage, Hypo LI, or β-score at two years. The remaining variables are the same as those observed at one-year after the initial and final transplants and will be analyzed in exactly the same way.

3.7.8 Sensitivity Analysis

The measure of a secondary endpoints may be missing at either 75±5 days or 365±14 days or both, following the initial and final infusion (if applicable). In our analysis report for the secondary endpoints, we will not impute missing values in any of these secondary endpoints. With close monitoring and data validation, the chance of having missing values will be minimized except that some serious adverse events (SAEs) may force subjects to withdraw from the study treatment or withdraw consent without a particular reason during the study. The interpretation of our analysis is strictly applied to the efficacy measurable population. However, a sensitivity analysis will be performed to determine the potential effect that missing values have on the analyses of the secondary endpoints.

- Missing values due to SAE will be excluded from this analysis. The analysis on these subjects will be provided in SAP section 3.8.
- We will conduct a sensitivity analysis assuming a missing at random mechanism: *i.e.*, we assume that a subject will withdraw consent at random with probability *r*. This probability will be estimated at 75 ± 5 days and 365 ± 14 days following the initial and final infusion (if applicable). The sensitivity analysis will be conducted according to the following two templates (the assumption of missing at random cannot be verified and these results must be viewed with caution.)

For the binary secondary endpoint under consideration, we will fit a logistic regression using the complete baseline data as described in SAP section 3.6. For each missing value, we will draw the outcome randomly from the Bernoulli distribution where the probability is that estimated from the corresponding complete data logistic regression model. The proportion of the secondary endpoint where missing values have been imputed will be compared to that based on the complete data to determine the magnitude of missing values' impact. If the difference between the proportion containing imputed values and that based on complete data is large, then the impact of missing values on the analysis of the endpoint under consideration is large as well; if the difference is small, then the impact of missing values on the analysis is small as well.

We will describe the reliability of this impact using the bootstrap procedure as described below:

1) We will resample the complete observations with replacement 1000 times.

2) For each bootstrap sample, we will estimate the proportion with missing values imputed in the way described above and calculate the difference between this proportion and that with the complete data.
3) We will calculate the standard deviation and construct the 95% bootstrap confidence interval based on the 1000 estimates of the difference.

For a continuous secondary endpoint, we will fit a linear regression using the complete data, given all the baseline information as described in SAP section 3.6. For each missing value, we will impute the outcome by the linear regression using the baseline data. The average of this secondary endpoint with missing value imputed this way will be compared to that based on the complete data to determine the magnitude of missing values' impact.

We will describe the reliability of this magnitude using the bootstrap procedure as described below:

1) We will resample the complete observations with replacement 1000 times.
2) For each bootstrap sample, we will estimate the average with missing values imputed in the way described above and calculate the difference between this average and that with the complete data.
3) We will calculate the standard deviation and construct the 95% bootstrap confidence interval based on the 1000 estimates of the difference.

At both 75±5 and 365±14 days following the initial and final islet transplant (if applicable), we will adopt this sensitive analysis using the methods described above.

3.8 *Safety Analyses*

Safety analyses will be conducted for the safety population. Summaries will be prepared for the targeted safety endpoints listed in the secondary endpoints and for all observed AEs organized by body system. The CIT consortium modified the National Cancer Institute (NCI) toxicity table to create a document relevant for trials of adult pancreatic islet transplantation. The resulting reference manual, "Terminology Criteria for Adverse Events (TCAE) In Trials of Adult Pancreatic Islet Transplantation," provides descriptive terminology and a grading (severity) scale which will be utilized for adverse event (AE) reporting.

Regular safety analyses will be prepared for the Data Safety Monitoring Board (DSMB). These summaries will be used to monitor the overall safety profile of the study. These analyses will summarize all AE data that are available at the time of the DSMB meeting. Analyses will summarize AEs by Medra term and body system. Separate incidence summaries will be prepared for serious AEs, for nonserious AEs and for all AEs combined. Separate tables will summarize severity and attribution. Each AE can be attributed to the investigational agent (allogeneic islets), the immunosuppression, both or neither. Identical safety summaries will be included in the final statistical report.

The protocol also describes targeted safety endpoints. The planned analyses for general safety outcomes and for these targeted safety endpoints are described in the following few sections.

3.8.1 Adverse Events

An AE is any occurrence or worsening of an undesirable or unintended sign, symptom (including an abnormal laboratory finding), or disease that is temporally associated with the use of a medicinal product whether considered related to the medicinal product or not. An SAE is defined as any AE occurring at any dose that suggests a significant hazard, contraindication, side effect, or precaution. This includes but is not limited to any of the following events (21CFR§312.32):

- Death.
- A life-threatening event. A life-threatening event is any adverse therapy experience that, in the view of the investigator, places the patient or participant at immediate risk of death from the reaction as it occurred.
- Inpatient hospitalization or prolongation of existing hospitalization.
- Persistent or significant disability.
- Congenital anomaly or birth defect.
- An event that required intervention to prevent permanent impairment or damage. An important medical event that may not result in death, be life threatening, or require hospitalization may be considered an SAE when, based on appropriate medical judgment, it may jeopardize the participant and may require medical or surgical intervention to prevent one of the outcomes listed above.
- Other conditions specified in the protocol.

In addition, events that occur at a higher than expected frequency, as determined by appropriate medical judgment, may be considered SAEs.

AEs will be graded on a scale from 1 to 5 according to the following standards in the *CIT-TCAE* manual:

Grade 1 = Mild adverse event.
Grade 2 = Moderate adverse event.
Grade 3 = Severe and undesirable adverse event.
Grade 4 = Life-threatening or disabling adverse event.
Grade 5 = Death.

AEs not included in the CIT-TCAE listing, will be recorded and graded 1 to 5 according to the General Grade Definition provided as in the table below:

Table 8-General Severity Definition of Adverse Event

Grade	Description	Definition
Grade 1	Mild	Transient or mild discomforts (<48 hours), no or minimal medical intervention/therapy required, hospitalization not necessary (non-prescription or single-use prescription therapy may be employed to relieve symptoms, e.g., aspirin for simple headache, acetaminophen for post surgical pain).
Grade 2	Moderate	Mild to moderate limitation in activity some assistance may be needed; no or minimal intervention/therapy required, hospitalization possible.
Grade 3	Severe	Marked limitation in activity, some assistance usually required; medical intervention/therapy required hospitalization possible.
Grade 4	Life-threatening	Extreme limitation in activity, significant assistance required; significant medical/therapy intervention required hospitalization or hospice care probable.
Grade 5	Death	Death.

All AEs will be reported and graded whether they are or are not related to disease progression or treatment. The relationship of an AE to islet transplantation, which includes the transplant procedure and/or the islet product, or to the immunosuppression and/or infection prophylaxis will be defined by using the descriptors provided in Table 9.

Table 9 -Attribution of Adverse Event

Code	Description	Definition
UNRELATED CATEGORY		
1	Unrelated	This adverse event is clearly not related to allogeneic islets; the islet transplant procedure; immunosuppression or infection prophylaxis.
RELATED CATEGORIES		
2	Unlikely	The adverse event is doubtfully related to allogeneic islets; the islet transplant procedure; immunosuppression or infection prophylaxis.
3	Possible	The adverse event may be related to allogeneic islets; the islet transplant procedure; immunosuppression or infection prophylaxis.
4	Probable	The adverse event is likely related to allogeneic islets; the islet transplant procedure; immunosuppression or infection prophylaxis.
5	Definite	The adverse event is clearly related to allogeneic islets; the islet transplant procedure; immunosuppression or infection prophylaxis.

The collecting and recording procedures for AEs are described in detail in Section 8.2 of CIT-07 protocol.

3.8.2 Analysis of Secondary Endpoints for Safety

The safety secondary endpoints in this study target AEs related to islet transplantation (transplant procedure and/or islet product) and the immunosuppression and/or infection prophylaxis.

The targeted AEs related to the islet transplantation include:

1. AE-IIP-1: Bleeding (>2 g/dL decrease in hemoglobin concentration)
2. AE-IIP-2: Segmental portal vein thrombosis
3. AE-IIP-3: Biliary puncture
4. AE-IIP-4: Wound complication (infection or subsequent hernia)
5. AE-IIP-5: Increased transaminase levels (>5 times ULN)

The targeted AEs related to the immunosuppression and infection prophylaxis therapy include:

1. AE-IP-1: Allergy
2. AE-IP-2: Reduction in GFR
3. AE-IP-3: Increase in urinary albumin excretion
4. AE-IP-4: Addition or intensification of anti-hypertensive therapy
5. AE-IP-5: Addition or intensification of anti-hyperlipidemic therapy
6. AE-IP-6: Oral ulcers
7. AE-IP-7: Lower extremity edema
8. AE-IP-8: Gastrointestinal toxicity
9. AE-IP-9: Neutropenia, Anemia, or Thrombocytopenia
10. AE-IP-10: Viral, Bacterial, or Fungal Infections
11. AE-IP-11: Benign or Malignant Neoplasms

3.8.2.1 Analyses at 75±5 days following the initial and final infusion

At 75±5 days following the initial and final islet transplant (if applicable), the incidence rates for each type of AE, grouped by severity, will be reported. For each type of AE, the number of resolved events will be counted as well. The mean, median, standard deviation and range for the number of days until the AE is resolved will be calculated. The results will be summarized in the table provided in SAP Appendix 2. Moreover, we will also report AEs categorized by attribution that is, AEs related to the islet transplant procedure or to immunosuppression; the report format is shown in SAP Appendix 3.

For second and third islet transplants, immunosuppression is modified by using Daclizumab (Zenapax®) instead of Thymoglobulin® for induction. For subjects who receive more than one islet transplant, we will compare the incidence rate at 75±5 days following the initial and final islet transplant for all AEs related to the islet transplant procedure and to immunosuppression one at a time using the McNemar's matched-pair test [8]. The claim of a difference in incidence rate will be made if the p-value is less

than 0.05. If the incidence is rare in certain severity categories for an AE, we will compare the incidence rate regardless the severity to accommodate the validity of the test.

The incidence rate of immune sensitization, defined as detection of anti-HLA antibodies at 75 ± 5 days following the initial and final islet transplant, (if applicable) will be reported. The exact two-sided 95% confidence interval of the incidence rate will also be reported.

3.8.2.2 Analyses at 365±14 days following the initial and final infusion

At 365±14 days following the initial and final islet transplant (if applicable), the incidence rates for each type of AE, grouped by severity, will be reported. For each type of AE, the number of resolved events will be counted as well. The mean, median, standard deviation and range for the number of days until the AE is resolved will be calculated. The results will be summarized in the table provided in Appendix 3. Moreover, we will also report the individuals who incur the AEs related to the islet transplant procedure and immunosuppression therapy, and the report is given in the table described in Appendix 3.

For subjects who require more than one islet transplant, we will compare the incidence rate at 365±14 days following the initial and final islet transplant for all the AEs related to the islet transplant procedure and immunosuppression therapy one at a time using the McNemar's matched-pair test [8]. The claim of difference in incidence rate will be made if the p-value is less than 0.05. If AE incidence is rare in certain severity categories, we will compare the incidence rate regardless of the severity to accommodate the validity of the test.

Adverse events may require a change in the immunosuppression for an individual transplant recipient. We will document the number and proportion of participants who require such a change.

For subjects who do not present the anti-HLA antibodies prior to transplantation, we will calculate the incidence rate of immune sensitization defined by detecting anti-HLA antibodies at 365±14 days following the initial and final islet transplant, (if applicable). The exact two-sided 95% confidence interval of the incidence rate will also be reported.

At 365±14 days following the initial islet transplant, we will also assess the change in retinal photography from pre-transplant. The incidence rate of worsening retinopathy will be calculated and the exact two-sided 95% confidence interval of the rate will be provided.

3.8.3 Statistical Analyses of Adverse Events

AEs will be analyzed according to the body system described in the *CIT-TCAE* manual. Any event that appears in a body system will be categorized into one of the five severity grades according to the *CIT-TCAE* manual or the general definition of severity given in Table 7 and analyzed accordingly. For each event, we will analyze its incidence rate per 100 person-days. Suppose for each subject, the incidence of a particular event is a Poisson process with the homogeneous incidence rate λ. At the time of analysis, if a subject is only followed r days after the randomization, the incidence of this event is distributed according to the Poisson distribution with mean $r\lambda/100$. The maximum

likelihood estimate of the incidence rate will be obtained based on data collected from the available subjects at the analysis time. The 95% confidence interval derived using the maximum likelihood estimator theory will be also reported. The results will be summarized in the table provided in SAP Appendix 4. Moreover, we will also list all the individuals who have ever had an incident of any of the AEs listed in Appendix 5 since their randomization.

3.9 *Quality of Life*

The analysis of QOL will be conducted for the ITT population.

3.9.1 Measurements of Quality of Life

Generic and disease-specific measures will be used to assess QOL. Questionnaires will be completed at enrollment and every 3 months during the screening period, then at day 75 and months 6 and 12 following transplant.

3.9.2 Generic Measure

3.9.2.1 Version 2 SF-36 survey

The version 2 SF-36 health survey, standard (4-week) recall form will be adopted for general QOL measure in this study. This widely used, generic instrument derives eight scales (physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, mental health) and two summary components (physical and mental).

3.9.2.2 European Quality of Life Questionnaire

The EQ-5D is a public domain instrument. This instrument is a utility measure that generates a descriptive profile and single index value for health status. The descriptive portion addresses five health dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). The second portion of the EQ-5D is a (0-100) visual analogue scale that is used to report overall health status.

3.9.3 Disease-Target Measures

3.9.3.1 Diabetes Distress Scale

The DDS represents the latest iteration of the Problem Areas in Diabetes (PAID) scale. This is a 17-item self-administered questionnaire selected from a longer battery of 28-items.

3.9.3.2 Hypoglycemia Fear Scale

The HFS is a 23-item self-administered survey for measuring the fear experienced with respect to hypoglycemia. The HFS has two subscales. The first measures hypoglycemia avoidance behavior, and the second measures worry about hypoglycemia.

3.9.4 Statistical Analyses of Quality of Life

The general analysis procedure is as follows:

1. We will estimate the mean QOL score and its 95% confidence interval at 75±5 days following the initial and final islet transplant.
2. We will estimate the mean QOL score and its 95% confidence interval at 365 ± 14 days following the initial and final islet transplant.
3. We will compare the QOL score at three time points: baseline, 75±5 days following the initial islet transplant and 365 ± 14 days following the initial islet transplant.

3.9.4.1 Analysis for SF-36 survey

3.9.4.1.1 *Analysis at 75±5 days following the initial and final islet transplant*

We will compute the Summary Physical Component (SPC) and Summary Mental Component (SMC) from the version 2 SF-36 health survey form at 75±5 days following the initial and final islet transplant (if applicable). For both SPC and SMC, we will calculate the sample mean and construct the 95% confidence interval of the mean value using the method for a continuous variable described in SAP section 3.6.3. We will also compute the change from baseline for both scores and analyze the changes using the same method.

3.9.4.1.2 *Analysis at 365±14 days following the initial and final islet transplant*

Obtain the SPC and SMC scores from the version 2 SF-36 health survey form at 365±14 days following the initial and final islet transplant (if applicable).

For both SPC and SMC, we calculate the sample mean and construct the 95% confidence interval of the mean value using the method for a continuous variable described in SAP Section 3.6.3.

3.9.4.1.3 *Comparison of SPC and SMC at three time points*

For each subject, we will obtain his/her score at baseline, 75±5 and 365±14 days following the initial islet transplant. The analysis adopts the method of linear mixed models with a term for time (baseline, 75 days following the initial islet transplant and 365 days following the initial islet transplant) and repeated measures on subjects.

The analysis proceeds as follows:

The Shapiro-Wilk test is applied to test the hypothesis that the QOL score is normally distributed. If the normality hypothesis is not rejected at significance level 0.05, a linear mixed model analysis will be performed to test whether the mean QOL score is the same among the three time points. This model will include a term for time (baseline, 75 days, and 365 days) with repeated measurements on subjects.

The results would be treated as follows:

1) If the p-value is greater than 0.05, we conclude there is no evidence to claim the difference of the QOL score from the baseline.

2) If the p-value is less than 0.05, we would compute the overall 95% confidence intervals for the pair-wise differences in the QOL score between any two time points using the Bonferroni adjustment for multiple tests.

If the normality hypothesis is rejected at significance level 0.05, we will use the bootstrap procedure (sampling subjects with replacement) to construct the overall 95% confidence interval of the pair-wise difference in the QOL score between any two time points using the Bonferroni adjustment for multiple tests.

We will use this to compare the difference at the three time points for both SPC and SMC. Any change in SPC or SMC from baseline may be due to the intervention or it may be due to regression to the mean. Because there is no control group, causality cannot be inferred.

3.9.4.2 Analysis for European Quality of Life Questionnaire

3.9.4.2.1 *Analysis at 75±5 days following the initial and final islet transplant*

Obtain the Overall Health Status from the EQ-5D instrument at 75±5 days following the initial and final islet transplant (if applicable). We calculate the sample mean and construct the 95% confidence interval of the mean OHS using the method for a continuous variable described in SAP section 3.6.3.

3.9.4.2.2 *Analysis at 365±14 days following the initial and final islet transplant*

Obtain the OHS from the EQ-5D instrument at 365±14 days following the initial and final islet transplant (if applicable). We calculate the sample mean and construct the 95% confidence interval of the mean OHS using the method for a continuous variable described in SAP section 3.6.3.

3.9.4.2.3 *Comparison of OHS at three time points*

For each subject, we obtain his/her OHS from the EQ-5D instrument at baseline, 75±5 and 365±14 days following the initial islet transplant. We compare the difference at the three time points for OHS using the linear mixed model method described in SAP section 3.9.4.1.3.

3.9.4.3 Analysis for Diabetes Distress Scale

3.9.4.3.1 *Analysis at 75±5 days following the initial and final islet transplant*

Obtain the DDS at 75±5 days following the initial and final islet transplant (if applicable). We calculate the sample mean and construct the 95% confidence interval of the mean DDS using the method for a continuous variable described in SAP section 3.6.3.

3.9.4.3.2 *Analysis at 365±14 days following the initial and final islet transplant*

Obtain the DDS at 354±14 days following the initial and final islet transplant (if applicable). We calculate the sample mean and construct the 95% confidence interval of the mean DDS using the method for a continuous variable described in SAP section 3.6.3.

3.9.4.3.3 *Comparison of DDS at three time points*

For each subject, we obtain his/her DDS at baseline, 75±5 and 365±14 days following the initial islet transplant. We compare the difference at the three time points for DDS using analysis of variance method described in SAP section 3.9.4.1.3.

3.9.4.4 Analysis for Hypoglycemia Fear Scale

3.9.4.4.1 *Analysis at75±5 days following the initial and final islet transplant*

Obtain the HFS at 75±5 days following the initial and final islet transplant (if applicable). We calculate the sample mean and construct the 95% confidence interval of the mean HFS using the method for a continuous variable described in SAP section 3.6.3.

3.9.4.4.2 *Analysis at 365±14 days following the initial and final islet transplant*

Obtain the HFS at 365±14 days following the initial and final islet transplant (if applicable). We calculate the sample mean and construct the 95% confidence interval of the mean HFS using the method for a continuous variable described in SAP section 3.6.3.

3.9.4.4.3 *Comparison of HFS at three time points*

For each subject, we obtain his/her HFS at baseline, 75 ± 5and 365 ± 14 days following the initial islet transplant. We compare the difference at the three time points for HFS using the analysis of variance method described in SAP section 3.9.4.1.3.

4. Interim Analyses and Safety Monitoring Analyses

The DSMB will be convened to review safety and efficacy data following National Institutes of Health (NIH) policy. When requested, formal interim analyses to assess safety and efficacy will be performed. Formal interim analyses will include distributions of endpoints, biomarkers and AEs. Additional analyses may be requested by the DSMB.

4.1 *Interim Analysis for Early Stopping*

This is a relatively small study; therefore, there is a need to collect as much safety data as possible, we do not plan to stop early for efficacy. We will monitor for lack of efficacy; the stopping rule will be based on excluding favorable outcome rates less than or equal to 30%.

The following table provides information on the study's strategy for stopping when the favorable outcome rate is too low. We will use the Lan and Demets [12] error spending approach with the O'Brien-Fleming [13] spending function. These calculations are based on using the O'Brien-Fleming spending function to calculate boundary values for a one-sided test of the hypothesis that the proportion achieving favorable outcome is no lower than a selected minimum value. The calculations assumed a conservative 2.5% level for the overall type I error. The procedure recommends terminating enrollment when there is overwhelming evidence that the favorable outcome rate is unacceptably low. Table 10 provides stopping boundaries for 20%, 30%, and 40% and for 3 and 4 planned analyses. It displays the numbers of favorable outcomes that would result in concluding that the true favorable outcome rate is less than the minimally acceptable rate.

For example, for three analyses and if the lowest acceptable favorable outcome rate were 20%, then the rule could not recommend stopping at the first interim analysis. It would recommend stopping at the second interim analysis (after 32 subjects had completed their one-year follow up) if none of the 32 subjects experienced a favorable outcome. The study would conclude that the favorable outcome rate was less than 20% at the end of the trial if 4 or fewer subjects experienced a favorable outcome. If the lowest acceptable favorable outcome rate were 30%, then the rule would recommend stopping at the second interim analysis if 3 or fewer subjects experienced a favorable outcome. It would conclude that the true favorable outcome rate was less than 30% after 48 subjects had completed the study if 8 or fewer subjects experienced a favorable outcome.

Table 10-Stopping Boundaries for Unacceptable Low Favorable Outcome Rates

Number of Interim Analyses	Interim Analysis	Number of Subjects	Minimally Acceptable Favorable Outcome Rate		
			20%	30%	40%
			Number Experiencing Favorable Outcome To Recommend Stopping		
3	1	16	--	--	--
	2	32	0	≤3	≤6
	3	48	≤4	≤8	≤12
4	1	12	--	--	--
	2	24	--	0	≤2
	3	36	≤1	≤4	≤7
	4	48	≤3	≤7	≤12

The Lan and Demets spending function approach allows for scheduling interim analysis as needed or requested by NIH or the DSMB. We anticipate that we will need to perform 4 interim analyses. For 4 equally spaced analyses, the stopping rule would not recommend stopping until 24 subjects have completed their one-year follow-up. At that time the rule would recommend stopping if there have been no favorable outcomes. After 36 subjects have completed their one-year follow-up the rule would recommend stopping if fewer than 5 subjects have experienced a favorable outcome. The final efficacy analysis will be adjusted to reflect the "alpha that has been spent" for these interim analyses by adjusting the confidence coefficient for the one-sided exact confidence interval for the primary outcome.

4.2 *Safety Monitoring Analyses*

AEs and clinical outcomes are monitored closely in this study. To protect the safety of subjects, safety stopping rules for the protocol and individuals sites have been developed.

4.2.1 Protocol Suspension and Review

Criteria for protocol suspension and review are detailed in the CIT-07 protocol, section 6.2.1.

After the protocol is placed on hold, no additional transplants within the trial will be performed at any participating clinical site until the CIT Steering Committee and DSMB meet either in person or by conference call to review in depth the results and circumstances surrounding the islet functional failure or SAE to determine whether the trial enrollment of new subjects and conduct of additional transplants could be safely resumed.

4.2.2 Site Suspension and Review

Criteria for suspension of study enrollment and initial islet infusions at individual sites are detailed in the CIT-07 protocol, section 6.2.2.

After any site is placed on hold, no additional transplants will be performed at that site until the CIT Steering Committee and DSMB meet either in person or by conference call to review in depth the results and circumstances surrounding the islet functional failure or SAE to determine whether the trial enrollment of new subjects and conduct of additional transplants could be safely resumed at that site, or whether there could be implications for the continuation of the entire proposed pilot protocol also at other affiliated sites testing the same protocol.

In all cases of PNF, subjects will be asked to temporarily continue immunosuppression to decrease the risk of sensitization that could increase the risk of poor outcome should future transplants occur. A tapering schedule will be applied until immunosuppressants are completely discontinued.

Reference

[1] Markmann JF, Deng S, Huang X, et al. Insulin independence following isolated islet transplantation and single islet infusions. *Ann Surg.* 2003; 237(6): 741-749; discussion 749-750.
[2] Service FJ, Molnar GD, Rosevear JW, Ackerman E, Gatewood LC, Taylor WF. Mean amplitude of glycemic excursions, a measure of diabetic instability. *Diabetes*. 1970; 19(9): 644-655.
[3] Ryan EA, Shandro T, Green K, et al. Assessment of the severity of hypoglycemia and glycemic lability in type 1 diabetic subjects undergoing islet transplantation. *Diabetes*. 2004; 53(4): 955-962.
[4] Ryan EA, Paty BW, Senior PA, Lakey JR, Bigam D, Shapiro AM. Beta-score: an assessment of beta-cell function after islet transplantation. *Diabetes Care*. 2005; 28(2):343-347.
[5] Saad MF, Anderson RL, Laws A, et al. A comparison between the minimal model and the glucose clamp in the assessment of insulin sensitivity across the spectrum of glucose tolerance. Insulin Resistance Atherosclerosis Study. *Diabetes*. 1994; 43(9): 1114-1121.
[6] Kessler L, Passermard R, Oberholzer J, et al. Reduction of blood glucose variability in type 1 diabetic patients treated by pancreatic islet transplantation: interest of continuous glucose monitoring. *Diabetes Care*. 2004; 27(3): 722-726.
[7] Steja D. Can continuous glucose monitoring provide objective documentation of hypoglycemia unawareness? *Endocr Pract*. 2005; 11(2):83-90.
[8] Rosner B. Fundamentals of Biostatistics, 6th Ed. *Duxbury*. 2006.
[9] Royston JP. Algorithm AS181: The W-test for normality. *Appl. Statist.* 1982; 31: 176-180.
[10] Zavaroni I, Deferrari G, Lugari R, et al. Renal metabolism of C-peptide in man. *J Clin Endocrinol Metab*. 1987; 65(3): 494-498.
[11] Zerbini G, Mangili R, Luzi L. Higher post-absorptive C-peptide levels in Type 1 diabetic patients without renal complications. *Diabetes Med.* 1999; 16(12): 1048.
[12] Lan KKG, DeMets DL. Changing frequency of interim analysis in sequential monitoring. *Biometrics*. 1989; 45:1017-1020.
[13] O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979; 35:549-556.
[14] Clarke WL, Cox DJ, Gonder-Frederik LA, Julian D, Schlundt D, Polonsky W. Reduced awareness of hypoglycemia in adults with IDDA. A prospective study of hypoglycemic frequency and associated symptoms. *Diabetes Care***.** 1995; 18(4): 517-522
[15] Spiegelhalter D, Thomas A, Best N, Lunn D. *WinBUGS User Manual*, 2003, www.mrc-bsu.cam.ac.uk/bugs
[16] R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. (ISBN 3-900051-07-0). http://www.R-project.org
[17] OpenBUGS. http://mathstat.helsinki.fi/openbugs/Home.html
[18] DeGroot, M.H. Optimal Statistical Decisions, McGraw Hill, New York, 1970.

[19] Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague ? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Statistics in Medicine,2005; 24:2401-2428.

[20] Benjamini Y, Hochberg Y. Controllign the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological), 57:289-300.

5. Appendices

Appendix 1 - Descriptive Statistics of Baseline Data

	Center						
	Univ Of Alberta	**Univ of Miami**	**Univ of Minnesota**	**Univ Of Pennsylvania**	**Emory Univ**	**Northwestern Univ**	**Total**
Demographic Variables							
n mean s.d. median range							
Sex: n (%) 1. male 2. female 3. total							
Race: n (%) 1. White 2. Black 3. Hispanic 4. Asian 5. Other							
Disease Factors							
Insulin Req. n mean s.d. median range							
HbA1c n mean s.d. median range							
# of Severe Hypo. n mean s.d. median range							
MAGE n mean s.d. median							

range							
LI n mean s.d. median range							
Clarke Score n mean s.d. median range							
HYPO score n mean s.d. median range							
β-score n mean s.d. median range							
C-peptide glucose Creatinine Ratio n mean s.d. median range							
Physical Conditions							
Body Weight n mean s.d. median range							
Height n mean s.d. median range							
QOL 1. SPC n mean s.d. median range 2. SMC n							

mean s.d. median range 3. OHS n mean s.d. median range 4. DDS n mean s.d. median range 5. HFS n mean s.d. median range							

Appendix 2 - Summary Table of Adverse Events Related to the Islet transplant and Immunosuppression Therapy

	Degrees of Event Severity					
Events	**Grade 1**	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**	**Total**
AE related to the islet transplantation (islet product or islet transplant procedure):						
AE-IIP-1 Number (%) Resolved? Number (%) mean s.d. median min-max						
AE-IIP-2 Number (%) Resolved? Number (%) mean s.d. median min-max						
AE-IIP-3 Number (%) Resolved? Number (%) mean s.d. median min-max						
AE-IIP-4 Number (%) Resolved? Number (%) mean s.d. median min-max						
AE-IIP-5 Number (%) Resolved? Number (%) mean s.d. median min-max						
AE related to the immunosuppresion and/or infection prophylaxis:						
AE-IP-1 Number (%) Resolved? Number (%) mean s.d.						

median						
min-max						
AE-IP-2						
Number (%)						
Resolved?						
Number (%)						
mean						
s.d.						
median						
min-max						
AE-IP-3						
Number (%)						
Resolved?						
Number (%)						
mean						
s.d.						
median						
min-max						
AE-IP-4						
Number (%)						
Resolved?						
Number (%)						
mean						
s.d.						
median						
min-max						
AE-IP-5						
Number (%)						
Resolved?						
Number (%)						
mean						
s.d.						
median						
min-max						
AE-IP-6						
Number (%)						
Resolved?						
Number (%)						
mean						
s.d.						
median						
min-max						
AE-IP-7						
Number (%)						
Resolved?						
Number (%)						
mean						
s.d.						
median						
min-max						
AE-IP-8						
Number (%)						
Resolved?						
Number (%)						

mean s.d. median min-max **AE-IP-9** Number (%) Resolved? Number (%) mean s.d. median min-max **AE-IP-10** Number (%) Resolved? Number (%) mean s.d. median min-max **AE-IP-11** Number (%) Resolved? Number (%) mean s.d. median min-max						

Appendix 3 - Adverse Events: Number Observed and Rate with Patient Identifications Grouped by Severity and Attribution

	Degrees of Event Severity					
Events	**Grade 1**	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**	**Total**
AE related to the islet transplantation (islet product or islet transplant procedure):						
AE-IIP-1 incidences	6 (12.5%) C11+ C12 C22 C34 C63 C66					
AE-IIP-2 incidences						
AE-IIP-3 incidences						
AE-IIP-4 incidences						
AE-IIP-5 incidences						
AE related to the immunosuppression and/or infection prophylaxis:						
AE-IP-1 incidences						
AE-IP-2 incidences						
AE-IP-3 incidences						
AE-IP-4 incidences						
AE-IP-5 incidences						
AE-IP-6						

incidences **AE-IP-7** incidences **AE-IP-8** incidences **AE-IP-9** incidences **AE-IP-10** incidences **AE-IP-11** incidences						

+ : Patient's identification, for example C11 stands for the incidence occurring on patient number 1 at Center #1

Appendix 4 - Analysis of Adverse Events at Month # since the Randomization

	AE related to the islet transplantation (islet product or islet transplant procedure) or immunosuppression and/or infection prophylaxis			AE **not** related to the islet transplantation (islet product or islet transplant procedure) or immunosuppression and/or infection prophylaxis		
All Adverse Events	**# of incidences**	**The incidence rate per 100 person-days (λ)**	**The 95 % confidence interval of λ based on MLE theory**	**# of incidences**	**The incidence rate per 100 person-days (λ)**	**The 95 % confidence interval of λ based on MLE theory**
Body System A **1. Grade 5** **2. Grade 4** Event 1 Event 2 ⋮ **3. Grade 3** Event 1 Event 2 ⋮ **4. Grade 2** Event 1 Event 2 ⋮ **5. Grade 1** Event 1 Event 2 ⋮						
Body System B **1. Grade 5** **2. Grade 4** Event 1 Event 2 ⋮ **3. Grade 3** Event 1 Event 2 ⋮ **4. Grade 2** Event 1 Event 2 ⋮ **5. Grade 1** Event 1						

Event 2 ⋮						
Body System C **1. Grade 5** **2. Grade 4** Event 1 Event 2 ⋮ **3. Grade 3** Event 1 Event 2 ⋮ **4. Grade 2** Event 1 Event 2 ⋮ **5. Grade 1** Event 1 Event 2 ⋮						
⋮						

Appendix 5 - Identification of Adverse Events at Month # since the Randomization

	AE related to the islet transplantation (islet product or islet transplant procedure) or immunosuppression and/or infection prophylaxis		AE **not** related to the islet transplantation (islet product or islet transplant procedure) or immunosuppression and/or infection prophylaxis	
All Adverse Events	**Identification of incidences**	**Number and percentage of patients who have had the incidence**	**Identification of incidences**	**Number and percentage of patients who have had the incidence**
Body System A **1. Grade 5** **2. Grade 4** Event 1 Event 2 ⋮ **3. Grade 3** Event 1 Event 2 ⋮ **4. Grade 2** Event 1 Event 2 ⋮ **5. Grade 1** Event 1 Event 2 ⋮	C111 C112 C113* C221 C231 C232 C331 C421 C461	6 (12.5%)		
Body System B **1. Grade 5** **2. Grade 4** Event 1 Event 2 ⋮ **3. Grade 3** Event 1 Event 2 ⋮ **4. Grade 2** Event 1 Event 2 ⋮				

5. Grade 1 Event 1 Event 2 ⋮				
Body System C **1. Grade 5** **2. Grade 4** Event 1 Event 2 ⋮ **3. Grade 3** Event 1 Event 2 ⋮ **4. Grade 2** Event 1 Event 2 ⋮ **5. Grade 1** Event 1 Event 2 ⋮				
⋮				

***: C113 stands for the third incidence of Event 1 of Grade 4 in body system A that occurs in patient #1 at Center 1.**

Appendix 6 - Summary of Protocol Deviations

	Number of Protocol Deviations					
Centers	**PD1**	**PD2**	**PD3**	**PD4**	**PD5**	**Total**
University of Alberta						
University of Miami						
University of Minnesota						
University of Pennsylvania						
Emory University						
Northwestern University						
University of California San Francisco						
University of Illinois Chicago						

Appendix 7 - Summary of excluded subjects

Centers	Number of excluded subjects	Reason for exclusion* (n)	# of excluded / (# of excluded + # of included)
University of Alberta			
University of Miami			
University of Minnesota			
University of Pennsylvania			
Emory University			
Northwestern University			
University of California San Francisco			
University of Illinois Chicago			

***** From list of 29 possible reasons.

Appendix 8 – Description of Bayesian hierarchical model for subgroup analysis and summary of MCMC based sensitivity analyses

Appendix 8.1: Bayesian Random-Effects Model for CIT-07

This appendix was prepared for the CIT Steering Committee by Kathryn Chaloner PhD, M. Kathryn Cowles PhD, and PhD students Emine Bayman, Qian Shi and Lixun Zhang. It contains the results of computationally intensive simulations to evaluate the Bayesian subgroup analysis and the proposed criteria for evaluating individual sites given in section 3.7.2.1.. The analysis involves MCMC to analyze the final data, and this report summarizes the results of simulating the MCMC analysis for several scenarios of underlying probabilities.

1. Background

The trial data will be analyzed using the WinBUGS program [1]. The hypothetical data sets in this report were also analyzed using WinBUGS. Unless stated otherwise the simulations for this report were generated using the R program [3] with OpenBUGS [4] with the BRugs interface to call OpenBUGS from R. In a few cases (involving extreme cases where individual centers may have results of 100% or 0% favorable outcomes) R2WinBUGS was used to avoid some complications where the sampler failed to satisfy convergence diagnostics, in which case those cases were not included in the calculations (less than 2% of the simulations).

2. The model and notation

The model and notation are now summarized. Suppose there are k centers and each center recruits n_i subjects, $i=1,..,k$: it is anticipated that $k=6$ and $(n_1, n_2, n_3, n_4, n_5, n_6) = (12, 12, 6, 6, 6, 6)$, or values close to these. Let y_i denote the observed number of favorable outcomes at center i out of n_i subjects, and let p_i denote the true underlying success probability at the ith center. The observed number of successes for each center, y_i, conditional on p_i, is a draw from a binomial distribution, independently at each center:

$y_i|p_i$ is Binomial(n_i, p_i)

for $i=1,...,k$. Further denote the log odds of success at each center by θ_i

$\theta_i = log[p_i/(1-p_i)]$

for $i=1,...,k$. Furthermore, assume that there may be between-center variability, but all centers are exchangeable: let θ_i given μ and σ^2 be like a sample from a normal population:

$\theta|\ \mu, \sigma^2$ is N(μ, σ^2).

The values μ and σ represent the population log odds and population between-center standard deviation of the log-odds respectively.

Let $\pi=e^{\mu}/(1+e^{\mu})$, the inverse log odds transformation, then π is the population probability of favorable outcome, the underlying predicted success probability for the population of centers.

A Bayesian approach, using MCMC methods, will be implemented to estimate parameters. The parameters of direct interest are: the overall probability of favorable outcome π, and each of the probabilities of favorable outcome at individual centers p_i, $i=1,\ldots,k$.

The Prior Distribution

The analysis of the CIT-07 data uses a very vague prior distribution with the distribution of μ assumed uniform and the distribution of σ^{-2}, denoted as τ, assumed to have an independent and proper gamma distribution with a large standard deviation. Specifically, τ has a density proportional to $\tau^{\alpha-1} e^{-\beta\tau}$ for $\tau>0$ [5, p. 39]: values of $\alpha=2$ and $\beta=1.5$ are specified for analysis. These values lead to stable estimation and are consistent with having very little prior information on the between center variability [6]. The location of the prior mean of the precision is also approximately consistent with historical data and the variance has been inflated to represent uncertainty.

The gamma prior distribution for τ has a mean of α/β and variance of α/β^2. The distribution of the θ_i's given μ is a scaled t-distribution [5, p. 42]. If, for example μ corresponds to a mean for all centers of $\pi = 0.7$, then the conditional prior distribution of any of the p_i (that is the distribution of $p_i|\mu$ is such that it lies between 0.17 and 0.96 with probability 0.95.

An additional setting of $\alpha=2$ and $\beta=0.75$ is examined in detail in this document. This gives a smaller prior mean value for the between center variability, which typically leads to more shrinkage towards the overall mean.

Proposed Evaluation of Centers

Twelve subjects at one center is considered sufficient to demonstrate comparable efficacy at that center. This is based on the power calculations given in this Appendix.

Two methods for the estimation of the rate at individual centers are used: 1) estimation based on the complete set of endpoint data from CIT-07 centers using data from CIT protocols 02, 03, 04, 05, 06, and 07 and 2) estimation based on the endpoint data from CIT-07 centers using data from CIT-07 alone.

Estimation of the rate at individual centers based on the complete set of endpoint data from CIT-07 centers using data from CIT protocols 02, 03, 04, 05, 06, and 07

Those CIT-02, -03, -04, -05, -06, and -07 subjects enrolled at a CIT-07 participating center are assessed for meeting the CIT-07 primary endpoint definition. If a center enrolls at least

12 subjects from the combination of the CIT-02, -03, -04, -05, -06, and -07 protocols, the criteria for evaluations of that center require that the resulting analysis satisfy both (a) and (b) below:

a) The overall primary result of the combined protocols is positive: that is, the primary analysis, which constructs a one sided 95% confidence interval for the overall probability of favorable outcome assuming no between center variability, rules out a favorable outcome rate of 0.50 or less.

b) The favorable outcome rate at that center, under this Bayesian model, is estimated to be at least 0.45 with posterior probability 0.90. That is, the lower 90% probability bound is at least 0.45.

Estimation of the rate at individual centers based on the endpoint data from CIT-07 centers using data from CIT-07 alone.

The subgroup analysis described in SAP section 3.7.2.1.1 will also be fit to a dataset comprised of the CIT-07 subjects alone. If a center enrolls at least 12 subjects in CIT-07, the criteria for evaluations of that center require that the resulting analysis satisfy both (a) and (b) below:

a) The overall primary result of CIT-07 is positive: that is, the primary analysis, which constructs a one-sided 95% confidence interval for the overall probability of favorable outcome assuming no between center variability, rules out a favorable outcome rate of 0.50 or less.

b) The favorable outcome rate at that center, under the Bayesian model, is estimated to be at least 0.45 with posterior probability 0.90. That is, the lower 90% probability bound is at least 0.45.

Power Calculations (α=2, β=1.5)

For any fixed value of the p_is, binomial data can be simulated and the results analyzed and examined as to whether the criteria above are met. This will give an estimate of "power" where power is defined as the probability of meeting the criteria under various hypothesized parameter values.

For this section we assume the sample sizes are two centers each with 12 subjects and 4 centers each with 6 subjects, for a total of n=48 subjects in CIT-07.

For illustration, suppose that $\pi = 0.70$ and there is no between center variability and so $p_1=p_2=p_3=p_4=p_5=p_6=0.70$. A total of 10,000 data sets were simulated and the proportion where criterion (a) was satisfied (the global power) was calculated as was the proportion where an individual center satisfied both criterion (a) and criterion (b) (the individual power). The individual power depends on the sample size in each group and the results are given in Table 5.

Additional Power Calculations

For completeness additional simulations were run assuming different true rates of favorable outcomes at each center. In all cases, $n1=n2=12$ and $n3=n4=n5=n6=6$ was used as the sample size: two centers enrolling 12 subjects and 4 centers enrolling 6 subjects each. Centers enrolling 6 subjects in CIT-07, and 6 in a different CIT protocol, will have an individual power slightly higher than a center enrolling 12 subjects in CIT-07 with the same true probability of favorable outcome, but no more than the Global Power.

In Table A1, P is defined as $P=(p_1, p_2, p_3, p_4, p_5, p_6)$.

Table A1: Additional Power Calculations for α=2 and β=1.5 (The final 4 entries in this table were calculated using R2WinBUGS)

N=(12,12,6,6,6,6)	Global Power	Individual Power (n=12)
P=(0.7,0.7,0.7,0.7,0.5,0.5)	0.59	0.56
P=(0.7,0.7,0.7,0.7,0.7,0.5)	0.72	0.68
P=(0.7,0.7,0.7,0.7,0.7,0.3)	0.60	0.56
P=(0.5,0.7,0.7,0.7,0.7,0.7)	0.60	0.31 (center 1) 0.57 (center 2)
P=(0.3,0.7,0.7,0.7,0.7,0.7)	0.29	0.03(center 1) 0.29(center 2)
P=(0.9,0.9,0.9,0.9,0.5,0.5)	1.00	1.00
P=(0.9,0.9,0.9,0.9,0.9,0.5)	1.00	1.00
P=(0.5,0.9,0.9,0.9,0.9,0.9)	1.00	0.40 (center 1) 1.00 (center 2)
P=(0.3,0.9,0.9,0.9,0.9,0.9)	0.99	0.05 (center 1) 0.99 (center 2)

Hypothetical data sets for CIT-07 (α=2 and β=1.5)
Several hypothetical data sets have been analyzed for illustration.

In all examples to follow it is assumed that 12 subjects enroll in centers 1 and 2, and 6 in each of centers 3, 4, 5, 6. A data set where the overall success rate corresponds to 36 successes out of 48, corresponding to an unadjusted estimate of 0.75 is first presented. The table below gives estimates and the lower 90% bound for each center.

Example 1: (11,11,5,2,5,2) successes out of (12,12,6,6,6,6) , for α=2, β=1.5

	mean	Lower 90% bound
π	0.74	0.60
p_1	0.85	0.74
p_2	0.85	0.74
p_3	0.78	0.61
p_4	0.51	0.28
p_5	0.78	0.62
p_6	0.51	0.28

In this example, the 2 centers with low rates of favorable outcome do not impact the ability of the other centers, with higher rates, to meet the criteria.

Example 2: (11,10,3,3,3,3) successes out of (12,12,6,6,6,6), for α=2, β=1.5

	mean	Lower 90% bound
π	0.66	0.52
p_1	0.82	0.69
p_2	0.77	0.63
p_3	0.58	0.38
p_4	0.58	0.38
p_5	0.58	0.38
p_6	0.58	0.38

In the example above it again holds that centers with low success rates do not negatively affect the centers with high success rates meeting the criteria. In the example below, all centers meet the criteria, even the one with 50% favorable outcome rate that does not meet the criteria in the example above. If the favorable outcome rate is high enough at other centers, a center with only 50% favorable outcome rate will meet the criteria.

Example 3: (11,10,5,5,4,3) successes out of (12,12,6,6,6,6), for α=2, β=1.5

	mean	Lower 90% bound
π	0.79	0.68
p_1	0.85	0.75
$p2$	0.81	0.69
p_3	0.80	0.65
p_4	0.80	0.65
$p5$	0.73	0.56
p_6	0.66	0.46

Alternative Prior Distribution Gamma(α=2, β=0.75)
Additional simulations were run assuming different true rates of favorable outcomes at each center for an inverse gamma prior distribution on the between center variance with (α=2, β=0.75).

In all cases, $n_1=n_2=12$ and $n_3=n_4=n_5=n_6=6$: two centers enrolling 12 subjects and 4 centers enrolling 6 subjects. Again, enters enrolling 6 subjects in CIT-07 and 6 in a different CIT study will have an individual power slightly higher than a corresponding center enrolling 12 subjects in CIT-07, but no more than the Global Power.

Table A2 Power and Size for Alternative Prior Distribution (α=2, β=0.75)

Probability of Favorable Outcome at Each Center	Global Power	Individual Power (n=12)
0.5	0.03	0.026
0.7	0.84	0.79
0.8	1.00	0.99
0.9	1.00	1.00

Note that the power and size in A2 are typically slightly higher, than in Table 5 of section 3.7.2.1 but not much higher, consistent with the analysis being robust and stable. In a few cases the probability of a center with a rate of favorable outcome 0.5 or less has a higher probability of meeting the criteria, supporting the choice of β=0.75 for the analysis. Additional calculations are given in Table A3 below.

Table A3: Additional Power Calculations

N=(12,12,6,6,6,6)	Global Power	Individual Power (n=12)
P=(0.7,0.7,0.7,0.7,0.5,0.5)	0.59	0.57
P=(0.7,0.7,0.7,0.7,0.7,0.5)	0.72	0.68
P=(0.7,0.7,0.7,0.7,0.7,0.3)	0.59	0.56
P=(0.5,0.7,0.7,0.7,0.7,0.7)	0.59	0.37 (center 1) 0.57 (center 2)
P=(0.3,0.7,0.7,0.7,0.7,0.7)	0.30	0.05 (center 1) 0.29 (center 2)
P=(0.9,0.9,0.9,0.9,0.5,0.5)	1.00	1.00
P=(0.9,0.9,0.9,0.9,0.9,0.5)	1.00	1.00
P=(0.5,0.9,0.9,0.9,0.9,0.9)	1.00	0.58 (center 1) 1.00 (center 2)
P=(0.3,0.9,0.9,0.9,0.9,0.9)	0.99	0.10 (center 1) 0.99 (center 2)

Hypothetical data sets for CIT-07 for α=2, β=0.75
Several hypothetical data sets have been analyzed for illustration.

The table below gives estimates and the lower 90% for the rate of favorable outcome using the alternative prior distribution α=2, β=0.75.

In conclusion it is recommended that for the analysis uses the prior distribution of β=1.5. A sensitivity analysis to this prior distribution will also be performed on the data from CIT 07.

Example: (11,11,5,2,5,2) successes out of (12,12,6,6,6,6), for α=2, β=0.75

	mean	Lower 90% bound
π	0.74	0.61
p_1	0.84	0.72
p_2	0.83	0.71
p_3	0.77	0.61
p_4	0.55	0.33
p_5	0.77	0.62
p_6	0.55	0.33

Example: (11,10,3,3,3,3) successes out of (12,12,6,6,6,6), for α=2, β=0.75

	mean	Lower 90% bound
π	0.67	0.54
p_1	0.79	0.67
p_2	0.75	0.62
p_3	0.60	0.42
p_4	0.60	0.42
p_5	0.60	0.42
p_6	0.60	0.42

Example: (11,10,5,5,4,3) successes out of (12,12,6,6,6,6), for α=2, β=0.75

	mean	Lower 90% bound
π	0.79	0.69
p_1	0.84	0.74
p2	0.81	0.70
p_3	0.80	0.66
p_4	0.80	0.66
p5	0.75	0.60
p_6	0.70	0.52

Comment

The simulations have demonstrated that the planned Bayesian analysis has appropriate power and size properties: centers with a high rate of favorable outcome have an appropriately high probability of meeting the criteria and centers with a low rate of favorable outcome do not. The results of using the alternative prior distribution indicate that the prior distribution will result in stable estimation.

In general the advantage of using the Bayesian model is estimates for each center are more precise than just using the data for each center alone. This is because in the Bayesian model the data from all other centers are incorporated in the estimates for each center through the correlation structure.

References

1 Spiegelhalter D, Thomas A, Best N, Lunn D. *WinBUGS User Manual*, 2003, www.mrc-bsu.cam.ac.uk/bugs

2 Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, 2004. John Wiley & Sons, Chichester UK.

3 R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. (ISBN 3-900051-07-0). http://www.R-project.org

4 OpenBUGS. http://mathstat.helsinki.fi/openbugs/Home.html

5 DeGroot, M.H. *Optimal Statistical Decisions,* McGraw Hill, New York, 1970.

6 Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR, 2005. How vague is vague ? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine,* 24:2401-2428.

Interpretation of hyperparameters α and β

Technical Note:

1. The hyperparameters α and β can be interpreted as follows: recall that $\theta|\ \mu, \sigma^2$ is $N(\mu, \sigma^2)$, with a Gamma(α, β)distribution for $\sigma^{-2} = \tau$, the marginal distribution of $\theta|\mu$ is a scaled t-distribution (DeGroot, 1970, p. 42) with location parameter μ, 2α degrees of freedom and precision factor α/β (DeGroot, 1970, p. 170). Thus a 95% highest prior density region for θ_i, the log odds of a favorable outcome at any center, is $\mu \pm (\beta/\sqrt{\alpha})t_{.025,2\alpha}$ and a corresponding 95% interval for p_i can be calculated by transformation. For example:

    i. For α=2 and β=1.5 the 95% interval corresponds to an interval for the log odds of μ ± 2.945. If μ is the log odds of 0.70, then the 95% interval for the probability of favorable outcome corresponds to 0.11 to 0.98.

    ii. For α=2 and β=0.75 and μ the log odds of 0.70, the 95% interval for the probability of a favorable outcome is 0.35 to 0.91.

   The two gamma prior distributions therefore represent considerable uncertainty in the prior distribution of the θ_i's and, correspondingly, of the p_i's. The prior distribution chosen for analysis (α=2, β=1.5) is very conservative in that it represents more between center variability than is present in historical data. The second prior distribution (α=2, β=0.75) indicates that the between center variability is smaller and therefore usually leads to more shrinkage. The differences, however, in power, size and analysis of hypothetical data sets are small.

Appendix 8.2: WinBUGS Codes

```
model

# N : number of centers
# yi : observed number of success at center i out of ni subjects transplanted
# pi: true underlying success probability at ith center
# thetai: log odds of success at center i
# mu: population log odds
# tausq: population between center precision

{
for (i in 1:N) {
      y[i] ~ dbin(p[i], n[i])
      logit(p[i])  <- theta[i] theta[i]
      ~ dnorm( mu, tausq )
}
mu ~ dflat()
tausq ~ dgamma( alpha, beta)
pi  <- 1 / ( 1 + exp( - mu) )
}

data
# list of different data values
list(y = c(11, 11, 5, 2, 5, 2), n = c(12, 12, 6, 6, 6, 6), N=6, alpha= 2, beta=1.5)
#list(y = c(11, 10, 3, 3, 3 ,3), n = c(12, 12, 6, 6, 6, 6), N=6, alpha= 2, beta=1.5)
#list(y = c(11, 10, 5, 5, 4 ,3), n = c(12, 12, 6, 6, 6, 6), N=6, alpha= 2, beta=1.5)
#list(y = c(11, 11, 5, 2, 5 ,2), n = c(12, 12, 6, 6, 6, 6), N=6, alpha= 2, beta=0.75)
#list(y = c(11, 10, 3, 3, 3 ,3), n = c(12, 12, 6, 6, 6, 6), N=6, alpha= 2, beta=0.75)
#list(y = c(11, 10, 5, 5, 4 ,3), n = c(12, 12, 6, 6, 6, 6), N=6, alpha= 2, beta=0.75)


initials
# list of 3 sets of initials
# 2.197=logit(0.9) = log(0.9/0.1)

list(mu = -2.197,  tausq=0.1)

list(mu = 0,  tausq=1)

list(mu = 2.197,  tausq=10
```

Appendix 8.3: R Program using BRugs Package, (alpha, beta)=(2, 0.75)

```
**************************************************************************
WinBUGS Code for 1-arm Study
**************************************************************************

# onearmislet_brugs.txt saved in working directory (I:/simulations/onearmislet_brugs.txt)

# N : number of centers
# yi : observed number of success at center i out of ni subjects transplanted
# pi: true underlying success probability at ith center
# thetai: log odds of success at center i
# mu: population log odds
# tausq: population between center precision
# alpha and beta: hyper parameters to specify Gamma distribution on between center #precision

model
{
for (i in 1:N) {
      y[i] ~ dbin(p[i], n[i])
      logit(p[i])  <- theta[i] theta[i]
      ~ dnorm( mu, tausq )
}
mu ~ dflat()
tausq ~ dgamma( alpha, beta)
pi  <- 1 / ( 1 + exp( - mu) )
}
```

```
**************************************************************************
This provides code for calling WinBUGS from R to simulate data multiple
times and store output in order to calculate power.

R Code to use BRugs library
**************************************************************************

working.directory <- "I:/simulations"
setwd(working.directory)

# p:     matrix of probabilities for each center
# reps: replications, number of data sets simulated
# seed: a seed for random numbers
```

```
brugs07 <- function(p, reps, seed, alpha, beta)
{

set.seed(seed)

# N : number of centers
N <- 6
library( BRugs )

invlogit <- function( l ) {
        1 / ( 1 + exp( -l ) )
}

logit <- function( p ) {
        log( p / (1-p) )
}

globallower <- rep(NA, reps)
indivlower <- matrix( NA, nrow = reps, ncol = 6)

# n: number of subjects within each center
n <- c( 12, 12, 6, 6, 6, 6 )
totn <- sum(n)

# theta: log odds of success at each center
theta <- logit( p )



for( rep in 1:reps ) {
        y <- rbinom( N, n, p )
        toty <- sum(y)

      globalresult <- binom.test(toty, totn, p = 0.6,
          alternative = "greater",
          conf.level = 0.95)  # 95% 1-sided c.i.

        globallower[rep] <- globalresult$conf.int[1]

        mydat <- list( n = n, y = y, N = N, alpha = alpha, beta = beta)

        inits1 <- list( mu = 0, tausq = 1 )
        myinits <- list(inits1)

        bugsData(mydat, file.path( getwd(), "data.txt"))
```

```
        bugsInits( myinits, numChains = 1, digits = 5 )
        modelCheck("onearmislet_brugs.txt")  # check model file
        modelData("data.txt")                # read data file
        modelCompile(numChains=1)                #compile model with 1 chain
        modelInits("inits1.txt")             #read initials data file
        modelGenInits()                          # generate initials
        modelUpdate(500)                         #burn in 500
        samplesSet(c("pi", "p"))             # pi and p should be
monitored
        modelUpdate(1500)                        #1500 more iterations

        # get the OpenBUGS simulated values of p's into an R matrix
        resultp <- matrix(NA, nrow=1500, ncol = N)
        for (j in 1:N) {
                nodename <- paste("p[",j,"]")
                resultp[,j] <- samplesSample(nodename)
        }

        # calculate the 0.1 quantile

if( !inherits(resultp, "try-error") )
        {
                indivlower[rep,] <- apply( resultp,2,quantile,0.1)
        }

# global temp stores number of times lower 95% CI bigger than 0.5
globaltemp <- as.numeric(globallower > 0.5)

# globalpower gives proportion of times lower 95% CI bigger than 0.5
globalpower <- sum(globaltemp, na.rm=TRUE) / reps

# indivtemp gives how many times 0.1 quantile bigger than 0.45
indivtemp <- matrix(as.numeric(indivlower > 0.45), nrow= reps )
for (i in 1:ncol(indivtemp) ) {
      indivtemp[,i] <- indivtemp[,i] * globaltemp
}

# goodreps: how many times we didn't see Na's in our simulation
goodreps <- length( indivlower[ !is.na(indivlower[,1]),1] )

indivpower <- apply(indivtemp,2,sum) / goodreps

print(rep)

}
cat ("  ----------------------------------","\n",
```

```
    " Simulation Results of Islet CIT07","\n",
    " ---------------------------------","\n",

    "\n", "--------------------------------------------------------------------------", "\n",
    " (alpha, beta)      :", "     ", alpha  ,   beta, "\n",
    " Sample Size       :", "     ", n, "\n",
    " Seed              :", "     ", seed, "\n",
    " p                 :", "     ", p, "\n",
   #" Sample logit      :", "     ", samplelogit, "\n",
   #" sdlogit           :", "     ", sdlogit, "\n",
    " Global Power      ", "     ", globalpower, "\n",
    " Individual Power:", "      ", indivpower, "\n",
    " # of Iterations    :", "     ", reps, "\n",
    " Goodreps          :", "     ", goodreps, "\n",
    " version           :", "     ", "BRugs" , "\n",

"--------------------------------------------------------------------------------", "\n")

}

#p <- rep(0.5, 6)
#p <- rep(0.7, 6)
#p <- c(0.7, 0.7, 0.7, 0.7, 0.5, 0.5)
#p <- c(0.7, 0.7, 0.7, 0.7, 0.7, 0.5)
#p <- c(0.7, 0.7, 0.7, 0.7, 0.7, 0.3)
#p <- c(0.5, 0.7, 0.7, 0.7, 0.7, 0.7)
p <- c(0.3, 0.7, 0.7, 0.7, 0.7, 0.7)

brugs07 (p, reps = 10000, seed = 278034, alpha=2, beta=0.75)
```

Appendix 8.4: R Program using R2WinBUGS Package, (alpha, beta)=(2, 0.75)

****
WinBUGS Code for 1-arm Study
****

```
#  onearmislet.txt, saved in "I:/r2wb/onearmislet.txt"

# N : number of centers
# yi : observed number of success at center i out of ni subjects transplanted
# pi: true underlying success probability at ith center
# thetai: log odds of success at center i
# mu: population log odds
# tausq: population between center precision



model
{
for (i in 1:N) {
       y[i] ~ dbin(p[i], n[i])
       logit(p[i])  <- theta[i] theta[i]
       ~ dnorm( mu, tausq )
}
mu ~ dflat()
tausq ~ dgamma( 2, 1.5)
pi  <- 1 / ( 1 + exp( - mu) )
}
```

****
This provides code for calling WinBUGS from R to simulate data multiple
times and store output in order to calculate power.

R Code to use R2WinBUGS library

****

```
#-------------------- onearmislet – R2WinBUGS-------------------------#

# p:      matrix of probabilities for each center
# reps: replications, number of data sets simulated
# seed: a seed for random numbers

isletCIT07 <- function(p, reps, seed)
{

set.seed(seed)

# N : number of centers
N <- 6
library( R2WinBUGS )

invlogit <- function( l ) {
        1 / ( 1 + exp( -l ) )
}

logit.fnc <- function(m) {
        log(m/(1-m))
}

globallower <- rep(NA, reps)
indivlower <- matrix( NA, nrow = reps, ncol = 6)

# n: number of subjects within each center
n <- c( 12, 12, 6, 6, 6, 6 )
totn <- sum(n)
samplelogit <- logit.fnc(p)

for( rep in 1:reps ) {

        # simulate a dataset

        y <- rbinom( N, n, p )
        toty <- sum(y)
# compute frequentist one-sided 95% confidence interval for overall success probability

      globalresult <- binom.test(toty, totn, p = 0.6, alternative = "greater",
            conf.level = 0.95)

# globallower stores    lower 95% confidence interval
```

```
globallower[rep] <- globalresult$conf.int[1]


        # set up input for WinBUGS

        mydat <- list( n = n, y = y, N = N)

        # give initial values, one set of initials.
        inits1 <- list( mu = 0, tausq = 1 )
        inits = list(inits1)

        # use "bugs" function in R2WinBUGS to invoke WinBUGS
        # monitor pi and p
        # number of chains= 2000, number of burn-in= 500
        results <- try( bugs( mydat, inits = inits, parameters.to.save =
        c("pi", "p"),
        model.file = "I:/r2wb/onearmislet.txt",
        DIC = FALSE,
        debug = FALSE,  n.chains = 1, n.iter=2000, n.burnin = 500,
        n.thin = 1) )


        # store 0.1 quantiles of each center's p if function didn't return an
error

if( !inherits(results, "try-error") )
        {
                # print(apply( results$sims.list$p,2,quantile,0.1))
                indivlower[rep,] <- apply(
results$sims.list$p,2,quantile,0.1)
        }

# global temp stores number of times lower 95% CI bigger than 0.5
globaltemp <- as.numeric(globallower > 0.5)

# globalpower gives proportion of times lower 95% CI bigger than 0.5
globalpower <- sum(globaltemp, na.rm=TRUE) / reps

# indivtemp gives how many times 0.1 quantile bigger than 0.45
indivtemp <- matrix(as.numeric(indivlower > 0.45), nrow= reps )
for (i in 1:ncol(indivtemp) ) {
      indivtemp[,i] <- indivtemp[,i] * globaltemp
}

# goodreps: how many times we didn't see Na's in our simulation
goodreps <- length( indivlower[ !is.na(indivlower[,1]),1] )
```

```
indivpower <- apply(indivtemp,2,sum) / goodreps

list(globalpower = globalpower, indivpower= indivpower)
}

# write simulation results to isletCIT07.csv file
column.name <- c("Global", "Center 1", "Center 2", "Center 3",
                 "Center 4", "Center 5", "Center 6")
res <- data.frame(cbind(globallower, indivlower))
names(res) <- column.name
write.table(res, file="C:/r2wb/isletCIT07.csv",
sep=",",
           col.names=TRUE,quote=FALSE, row.names=FALSE)

cat("  ----------------------------------","\n",
    " Simulation Results of Islet CIT07","\n",
    " ----------------------------------","\n",

    "\n", "-------------------------------------------------------------------------------", "\n",
    " Sample Size        :", "     ", n, "\n",
    " Seed               :", "     ", seed, "\n",
    " p                  :", "     ", p, "\n",
    " alpha, beta        :", "     ", "2 , 1.5", "\n",
    " Sample logit       :", "     ", samplelogit, "\n",
    " Global Power       :", "     ", globalpower, "\n",
    " Individual Power :", "     ", indivpower, "\n",
    " # of Iterations    :", "     ", reps, "\n",
    " Goodreps           :", "     ", goodreps, "\n",
    " Version            :", "     ", "R2WinBUGS", "\n",
    " Updater            :", "     ", "UpdaterSlice", "\n",

"-------------------------------------------------------------------------------", "\n")
}

#p <- rep(0.8, 6)
#p <- rep(0.9, 6)
#p <- c(0.9, 0.9, 0.9, 0.9, 0.5, 0.5)
#p <- c(0.9, 0.9, 0.9, 0.9, 0.9, 0.5)
#p <- c(0.5, 0.9, 0.9, 0.9, 0.9, 0.9)


p <- c(0.3, 0.9, 0.9, 0.9, 0.9, 0.9)

isletCIT07( p, reps=10000, seed=246299)
```

Appendix 9- Analysis Templates for the Secondary Efficacy Endpoints Measured at 75±5 Days Following the Initial and Final Infusion

	Center						
	Univ of Alberta	**Univ of Miami**	**Univ of Minnesota**	**Univ of Pennsylvania**	**Emory Univ**	**Northwestern Univ**	**Total**
Continuous Outcomes							
Insulin Req. n mean s.d. median range 95% CI of mean							
HbA1c n mean s.d. median range 95% CI of mean							
MAGE n mean s.d. median range 95% CI of mean							
LI n mean s.d. median range 95% CI of mean							
Hypo n mean s.d. median range 95% CI of mean							
Glucose Change n mean s.d. median range 95% CI of mean							
β-score n mean s.d.							

median range 95% CI of mean							
C-pep:glucose ratio n mean s.d. median range 95% CI of mean							
AIRglu n mean s.d. median range 95% CI of mean							
SI n mean s.d. median range 95% CI of mean							
DI n mean s.d. median range 95% CI of mean							
Glucose variability n mean s.d. median range 95% CI of mean							
# of Hypoglycemia n mean s.d. median range 95% CI of mean							
Duration of Hypoglycemia n mean s.d. median range							

95% CI of mean							
Binary Outcomes							
MAGE ≥ 11.1 mmol/l n proportion (p) 95% CI of p							
LI ≥ 433 mmol/l2/h·wk-1 n proportion (p) 95% CI of p							
Hypo ≥ 1047 n proportion (p) 95% CI of p							

Appendix 10- Analysis Templates for the Secondary Efficacy Endpoints Measured at 365±14 Days Following the Initial and Final Infusion

	Center						
	Univ of Alberta	**Univ of Miami**	**Univ of Minnesota**	**Univ of Pennsylvania**	**Emory Univ**	**Northwestern Univ**	**Total**
Continuous Outcomes							
Insulin Req. n mean s.d. median range 95% CI of mean							
HbA1c n mean s.d. median range 95% CI of mean							
MAGE n mean s.d. median range 95% CI of mean							
LI n mean s.d. median range 95% CI of mean							
Clarke Score n mean s.d. median range 95% CI of mean							
Hypo n mean s.d. median range 95% CI of mean							
Glucose Change n mean							

s.d. median range 95% CI of mean							
β-score n mean s.d. median range 95% CI of mean							
C-pep:glucose ratio n mean s.d. median range 95% CI of mean							
Glucose variability n mean s.d. median range 95% CI of mean							
Binary Outcomes							
MAGE ≥11.1 mmol/l n proportion (p) 95% CI of p							
LI ≥ 433 mmol/l2/h·wk-1 n proportion (p) 95% CI of p							
Hypo ≥ 1047 n proportion (p) 95% CI of p							
Insulin Indenp. n proportion (p) 95% CI of p							
Second infusion n proportion (p) 95% CI of p							
Third infusion n proportion (p) 95% CI of p							
HbA1c<6.5%							

and free of severe hypoglycemic events n proportion (p) 95% CI of p							