Attachment 9. Revised Power and Sample Size Calculations

The goal of the study is to conduct an epidemiologic, mixed-methods evaluation of MAT in real-world outpatient settings, incorporating characteristics for the site along with client-level covariates. The study is not powered to conduct site-level comparisons.

RTI used a conservative approach for the original power calculations presented in our initial response to the MAT Study RFP to determine the number of clients needed at the end of the 2-year observation period. Inputs to the initial power calculations included a 50 percent opioid abstinence rate for all four treatment arms. Though opioid abstinence rates may differ from 50 percent, this value produced a relatively large variance compared to other rates and a corresponding conservative increase in the required sample size. Additionally, we originally assumed descriptive instead of model-based statistics. Without the benefits of covariates to lower the model errors, our descriptive approach also produced slightly larger variances than what we would expect in our planned analyses.[1]

RTI has revisited the assumptions used in the original power calculations. Borrowing information from more recently published research and field experience to date, we determined that a smaller sample size than originally proposed would still allow us to meet the analytic needs for a successful MAT Study. Details of the revised power calculations are provided below.

Study retention rates were used to inflate the 24-month sample sizes obtained from the power calculation. This step accounts for loss from, for example, clients who drop out of the study or who we are unable to locate after completing the Baseline Client Questionnaire.[2] We estimated the values for the study retention rates based on prior experience from across the RTI project team. RTI assumes that the study retention rates will be relatively equal across the treatment conditions. This corresponds, for example, with results from the START Study (Hser et al. 2015), a multisite, randomized comparative study of MMT and BUP, where 73.6% MMT and 73.7% BUP study participants were recontacted 2-8 years after treatment initiation.

The MAT Study protocols used to retain clients follow those discussed in, for example, Lynn (2009, Chapters 10 and 12) and Desmond, et al. (1995); these researchers maintained a 98 percent retention rate after a one-year period, a level considered to unrealistic for studies of today. Additional details on the study retention protocols is found in the MAT Study *Plan to Recruit Sites and Study Participants*.

---

[1] Power calculations were tabulated for two arms using a conservative 50 percent abstinence rate to generate the largest variance.
[2] Study retention is associated only with participation in the MAT Study and not with retention in treatment. Participants are eligible for the MAT Study even if they drop out of OUD treatment entirely.

Power Calculation Results

Table 3-1 contains the number of primary study sites needed to fulfill the initial analytic objectives for the MAT Study outlined above. The result (i.e., 37 sites) was originally increased to allow for an even number of sites chosen by MSA for the MAT Study. To date, the MAT Study has exceeded the goal of 37 primary sites by 11 (48 primary site plus 4 feeder sites) and we did not revisit this portion of the power calculation.[3] Note that site recruitment efforts indicated that a balanced design was not feasible because of the differential site distribution across the MSAs.

Table 3-1. Minimum Number of MAT Study Sites by Site-level ICC and Effect Size

Site ICC	Standardized Difference in Treatment Rates[a]		
	0.1 (Small)	0.2	0.3 (Med)
0.001 (lower)	103	26	12
0.01 (low)	147	**37**	17
0.10 (high)	579	145	65

Note: ICC = intracluster correlation.
[a] The effect size is a standardized measure of the difference between two treatment rates relative to the pooled variance (Cohen 1988). Formula for the effect size and pooled variance are provided below in reference to equation (3.1).

Table 3-2 contains the number of clients per viable treatment arm required to power the study under the original (conservative) and revised intracluster correlation (ICC) assumptions and by number of analytic data points (i.e., 3 vs. 5 client interviews). The numbers correspond to those persons per treatment arm who participate in the full 2-year observation period and do not attrite from the MAT Study. Repeated-measures analyses using all 5 client interviews require many fewer participants than those involving baseline, 12- and 24-month interviews (e.g., 208 vs. 297 for 90% power and ICC=0.10).[4]

---

[3] RTI is currently recruiting sites in three new MSAs (Boston, Denver, and Raleigh/Durham) and anticipates adding to the study approximately four sites per MSA in February 2019.
[4] RTI and CDC held initial discussions about power prior to the inclusion of the 3- and 6-month Check-in interviews. Therefore, we include power calculations for both 3 and 5 data points (client interviews).

Table 3-2. Number of Clients per Treatment Arm Completing the 2-Year Study by Original and Revised Power Assumptions and Number of Interviews

Power Assumptions	Client ICC	2-Sided Test Statistic, 90% Power	
		3 Client Interviews	5 Client Interviews
Original (conservative)	0.40	**445** [a]	386
Revised	0.06	277	184
	0.10	**297** [b]	208

Note: ICC = intracluster correlation.
[a] Originally proposed sample size under conservative power calculation assumptions with 3 data points.
[b] Revised sample size under less conservative power calculation assumptions with 3 data points.

We approached the new client-level power calculations in a less conservative, more clinically appropriate manner. Changes were introduced based on review of recent literature and included:

- Decrease of the client-level intracluster correlation coefficient (ICC) from 0.40 to approximately 0.10 (Bachhuber et al. 2018; DeBar et al. 2018; Goodwin et al. 2018; Barnhart et al. 2016; Kuo et al. 2016; Allen et al. 2014; Thompson et al. 2012; Campbell et al. 2005).
- Increase in the percent of the variance explained in a model-based analyses from 0 (i.e., descriptive statistics only) to 0.20 (Heidebrecht et al. 2018; Levine et al. 2015).[5] See discussion of $\varepsilon = 0.20$ below for equation (3.2).
- Decrease of 0.10 in the covariance between two opioid abstinence rates to account for the many participating sites that provide multiple treatment regimens (i.e., arms). See discussion of $\rho = 0.10$ below for equation (3.2).

The proposed reduction of the client-level ICC noted above to clinically-feasible levels is worth further discussion. Two multisite studies found in the literature to have similar characteristics to the MAT Study are the START Study (Hser et al. 2015) and POATS (Weiss et al. 2015). The START Study included MMT and BUP in their evaluation (randomized to treatment within site), and POATS recruited for BUP OUD treatment only. Neither of the comparison studies included as many sites as targeted for the MAT Study—7 for START and 10 for POATS. Both journal articles discuss evaluating the client-level ICC for significance, and both concluded that the values were insignificant, resulting in a site-level fixed effect within their models.

As with the original power calculations in our initial proposal, we used a 50 percent opioid abstinence rate with the revised power calculations. Some studies suggest that this rate

---

[5] Candidate covariates include key health, quality of life, and socioeconomic outcomes as well as client- and provider-level characteristics such as sex and geolocation.

may be lower than 50 percent (see, e.g., Zhu et al. 2018; Bart 2012; Kleber 2007); values smaller than 50 percent would reduce the calculations shown in Table 3-2.

The results in Table 3-2 display the targeted number of clients in the MAT Study at the end of data collection (i.e., 24-month interview) per treatment arm. Because of some study loss (study withdrawal, loss to follow-up), the targeted number of clients was inflated by 50 percent based on experience from the RTI project team (Table 3-3). Therefore, the revised target number of enrolled clients across the four treatment arms is approximately 2,376 (=594 × 4).

Table 3-3. Target Number of Completed Baseline Interviews per Treatment Arm by Original and Revised Power Assumptions and Number of Interviews

| Power Assumptions | Client ICC | 2-Sided Test Statistic[a], 90% Power | |
		3 Client Interviews	5 Client Interviews
Original (conservative)[b]	0.40	**890**	772
Revised[c]	0.06	554	368
	0.10	**594**	416

Note: ICC = intracluster correlation.
[a] As discussed in previous meetings, attaining these numbers for NTX and COUN continues to be challenging. Recruitment results suggest that clients receiving either NTX or COUN may be fewer than originally hypothesized.
[b] Originally proposed sample size under conservative power calculation assumptions.
[c] Revised sample size under less conservative power calculation assumptions.

Note that the estimated number of baseline interviews (i.e., enrolled study participants) in Table 3-3 is larger than the MMT and BUP enrollment numbers for the START Study (Hser et al. 2015) and the BUP enrollment numbers for the POATS (Weiss et al. 2015).

Power Calculation Details

The formulae used to determine site and client sample sizes to satisfy the statistical power requirements are a function of several components. First, the intracluster correlation (ICC)

$$ICC = \frac{B^2}{B^2 + W^2}$$

measures the relative variation between site estimates (B^2) to the overall variance for that estimate ($B^2 + W^2$) where (W^2) is the "within site" variance component. The same is true for the homogeneity of estimate across clients in comparison to the change in their data examined across the 2-year observation period. Tables 3-1 and 3-2 above show variability in the minimum participating site and clients, respectively, for varying levels of ICC.

The formula used to determine the minimum number of study participants completing the 2-year observation period (n) is (Diggle et al. 2002; Hedeker et al. 1999):

$$n = 2\left[1 + (r - 1)ICC_p\right]\frac{(z_{\alpha/2} - z_{1-\beta})^2}{r\Delta^2}$$

<div align="right">(3.1)</div>

where r is the number of client interviews over the 2-year observation period, ICC_p is the assumed participant-level ICC, $z_{\alpha/2}$ =1.96 is the z-score associated with the two-sided test with alpha set to 0.05, $z_{1-\beta}$ is the z-score associated with power of the test (i.e., 1.28 for 90% power), and $\Delta^2 = \delta^2 / var(\delta)$ is the square of the specified detectable difference between two treatments, $\delta = p_1 - p_2$ =0.1 (10%), divided by the pooled variance:

$$var(\delta) = \left[s_1^2 + s_2^2 - 2\rho\sqrt{s_1^2 \times s_2^2}\right] \times (1 - \varepsilon)$$

<div align="right">(3.2)</div>

where $s_j = p_j(1 - p_j)$ for j=1,2; ρ=0.10, an estimated measure of association within site across different treatment arms; and ε=0.20, an estimate of the efficiency gained (i.e., reduction in variance) through a model-based analyses with strong covariates. Please see the accompanying Excel file, with additional citations, for the direct calculations.

For the calculations shown in **Table 3-2,** r=3 is the number of client interviews initially proposed (i.e., baseline, 12-month, and 24-month); ICC_p = 0.06, 0.1, and 0.4; and values of 0.5 (50% abstinence rate) and 0.4 were used for (generic) treatment group 1 and 2 for the detectable difference δ =0.1. We also used the treatment rates to calculate the pooled variance; $var(\delta)$ was reduced by 0.2 (i.e., model covariates could control 20% of the variability (Levine et al. 2015) and by a small correlation of 0.1 to account for clients within the same treatment facility. As shown in **Table 3-2** *and in* the accompanying Excel file, sample size requirements (3.1) are greatly reduced if data are analyzed using all r=5 client interviews.

The Excel spreadsheet accompanies this document for ease in understanding the calculations and for conducting additional sensitivity analyses through changes in the assumptions. For example, if reduce the client-level ICC below 0.6, then the resulting sample sizes are less than those shown in **Table 3-2**.

Considering the minimum number of clients, the total number of sites (m) was determined in combination with budgetary caps as

$$m = \frac{\nu}{\bar{n}C^2}\left[1 + ICC_s(\bar{n} - 1)\right]$$

<div align="right">(3.3)</div>

where $\bar{n}$ is the average number of clients expected at the end of the 2-year observation period, $\tilde{V}$ is the coefficient of variation specified in the RFTOP, C^2 is the square of the desired effect size (**Table 3-1**), and ICC_s is the site-level ICC values shown in **Table 3-2**.

Analysis of Subgroups within the MAT Study

Using (3.1) and the revised assumptions discussed above, we anticipate that any subgroup with approximately 300 participants completing the 24-month study period by treatment group is sufficient for a 10 percent detectable difference with at least 90 percent power.

References

Allen, K.D., Bosworth, H.B., Chatterjee, R., et al. (2014). Clinic variation in recruitment metrics, patient characteristics and treatment use in a randomized clinical trial of osteoarthritis management. BMC Musculoskeletal Disorders, 15:413. https://bmcmusculoskeletdisord.biomedcentral.com/articles/10.1186/1471-2474-15-413

Bachhuber, M.A., Nash, D., Southern, W.N., Heo, M., Berger, M., Schepis, M., and Cunningham, C.O. (2018). Reducing the default dispense quantity for new opioid analgesic prescriptions: study protocol for a cluster randomized controlled trial. BMJ Open, 8:e1-e7. https://bmjopen.bmj.com/content/bmjopen/8/4/e019559.full.pdf

Barnhart, D., Hertzmark, E., Liu, E., Mungure, E., Muya, A.N. et al. (2016). Intra-cluster correlation estimates for HIV-related outcomes from care and treatment clinics in Dar es Salaam, Tanzania. Contemporary Clinical Trials Communications, 4, 15 December 2016: 161-169. https://www.sciencedirect.com/science/article/pii/S2451865416300436

Bart, G. (2012). Maintenance Medication for Opiate Addiction: The Foundation of Recovery. J Addict Dis. 2012 Jul; 31(3): 207–225.

Campbell, M.K., Fayers, P.M., and Grimshaw, J.M. (2005). Determinants of the intracluster correlation coefficient in cluster randomized trials: the case of implementation research. Clin Trials. 2005;2(2):99-107. https://www.ncbi.nlm.nih.gov/pubmed/16279131

Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, New Jersey.

DeBar, L., Benes, L., Bonifay, A., et al. (2018). Interdisciplinary team-based care for patients with chronic pain on long-term opioid treatment in primary care (PPACT) – Protocol for a pragmatic cluster randomized trial. Contemporary Clinical Trials, 67, April 2018: 91-99. https://www.sciencedirect.com/science/article/pii/S1551714417305785

Desmond, D.P., Maddux, J.F., Johnson, T.H., and Confer, B.A. (1995). Obtaining Follow-Up Interviews for Treatment Evaluation. *Journal of Substance Abuse Treatment*, **12**(2): 95-102.

Diggle, P.J., Heagerty, P.J., Liang K-Y., and Zeger, S.L. (2002). *Analysis of longitudinal data,* 2nd edition. Oxford Statistical Science Series.

Goodwin, J.S., Kuo, Y-F., and Brown, D. (2018). Association of Chronic Opioid Use With Presidential Voting Patterns in US Counties in 2016. JAMA Netw Open. 2018;1(2):e180450. doi:10.1001/jamanetworkopen.2018.0450 https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2685627

Hedeker, D., Gibbons, R.D., and Waternaux, C. (1999). Sample size estimation for longitudinal designs with attrition. Journal of Educational and Behavioral Statistics, 24(1): 70-93.

Heidebrecht, F., MacLeod, M.B., and Dawkins, L. (2018). Predictors of heroin abstinence in opiate substitution therapy in heroin-only users and dual users of heroin and crack. Addictive Behaviors, 77:210–216. https://www.sciencedirect.com/science/article/pii/S0306460317303842?via%3Dihub

Hser, Y. Evans, E., Huang, D., Weiss, R., et al. (2015). Long-term outcomes after randomization to buprenorphine/naloxone versus methadone in a multi-site trial. Addiction, 111:695–705.

Kleber, H.D. (2007). Pharmacologic treatments for opioid dependence: detoxification and maintenance options. Dialogues Clin Neurosci. 2007 Dec; 9(4): 455–470.

Kuo, Y-F, Mukaila, A.R., Chen, N-W., Hasan, H., Goodwin, J.S. (2016). Trends in Opioid Prescriptions Among Part D Medicare Recipients From 2007 to 2012. The American Journal of Medicine, 129, 221.e21-221.e30. https://www.utmb.edu/pcor/App_Assets/assets/pubs/26522794.pdf

Levine, A.R., Lundahl, L.H., Ledgerwood, D.M., et al. (2015). Gender-Specific Predictors of Retention and Opioid Abstinence During Methadone Maintenance Treatment. Journal of Substance Abuse Treatment, 54: 37–43. https://www.sciencedirect.com/science/article/pii/S0740547215000240?via%3Dihub

Lynn, P., editor (2009). *Methodology of Longitudinal Surveys*. United Kingdom: John Wiley f& Sons, Ltd.

Thompson, D.M., Fernald, D.H., and Mold, J.W. (2012). Intraclass Correlation Coefficients Typical of Cluster-Randomized Studies: Estimates From the Robert Wood Johnson Prescription for Health Projects. Ann Fam Med. 2012 May; 10(3): 235–240. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3354973/

Weiss, R.D, Potter, J.S., Griffin, M.L., et al. (2015). Long-term Outcomes from the National Drug Abuse Treatment Clinical Trials Network Prescription Opioid Addiction

Treatment Study. Drug Alcohol Depend. 2015 May 1; 150: 112–119.
doi:10.1016/j.drugalcdep.2015.02.030.

Zhu, Y., Evans, E.A., Mooney, L.J., Saxon, A.J., Kelleghan, A., Yoo, C., and Hser, Y.I. (2018).
Correlates of Long-Term Opioid Abstinence After Randomization to Methadone Versus
Buprenorphine/Naloxone in a Multi-Site Trial. J Neuroimmune Pharmacol, 13(4):488-
497. doi: 10.1007/s11481-018-9801-x. Epub 2018 Aug 9.