Title

Impact of complex, partially nested clustering in a three-arm individually randomized group treatment trial: A case study with the wHOPE trial

Permalink

https://escholarship.org/uc/item/2kb021ww

Journal

Clinical Trials, 19(1)

ISSN

1740-7745

Authors

Tong, Guangyu
Seal, Karen H
Becker, William C
et al.

Publication Date

2022-02-01

DOI

10.1177/17407745211051288

Peer reviewed

Impact of complex, partially nested clustering in a three-arm individually randomized group treatment trial: A case study with the wHOPE trial

Guangyu Tong[1], **Karen H Seal**[2,3], **William C Becker**[4,5], **Fan Li**[1], **James D Dziura**[1,6], **Peter N Peduzzi**[1], **Denise A Esserman**[1]

[1]Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA

[2]San Francisco VA Health Care System, Integrative Health Service, San Francisco, CA, USA

[3]Department of Medicine and Psychiatry, University of California, San Francisco, CA, USA

[4]Yale School of Medicine, Department of Internal Medicine, New Haven, CT, USA

[5]VA Connecticut Healthcare System, Pain Research, Informatics, Multimoribidities and Education Center of Innovation, West Haven, CT, USA

[6]Yale School of Medicine, Department of Emergency Medicine, New Haven, CT, USA

Abstract

Background/Aims: When participants in individually randomized group treatment trials are treated by multiple clinicians or in multiple group treatment sessions throughout the trial, this induces partially nested clusters which can affect the power of a trial. We investigate this issue in the Whole Health Options and Pain Education trial (wHOPE), a three-arm pragmatic, individually randomized clinical trial. We evaluate whether partial clusters due to multiple visits delivered by different clinicians in the Whole Health Team arm and dynamic participant groups due to changing group leaders and/or participants across treatment sessions during treatment delivery in the Primary Care Group Education arm may impact the power of the trial. We also present a Bayesian approach to estimate the intraclass correlation coefficients (ICCs).

Methods: We present statistical models for each treatment arm of wHOPE in which power is estimated under different ICCs and mapping matrices between participants and clinicians or treatment sessions. Power calculations are based on pairwise comparisons. In practice, sample size calculations depend on estimates of the ICCs at the treatment sessions and clinician levels. To accommodate such complexities, we present a Bayesian framework for the estimation of ICCs under different participant-to-session and participant-to-clinician mapping scenarios. We simulated continuous outcome data based on various clinical scenarios in wHOPE using a range of ICCs and mapping matrices and used Gibbs samplers with conjugate priors to obtain posteriors of the ICCs

---

Corresponding Author: Guangyu Tong, Yale Center for Analytical Sciences, Suite 511, 300 George Street, New Haven, CT, USA 06511, guangyu.tong@yale.edu, Phone: (203) 785-4297.

under those different scenarios. Posterior means and medians and their biases are calculated for the ICCs to evaluate the operating characteristics of the Bayesian ICC estimators.

Results: Power for Whole Health Team vs. Primary Care Group Education is sensitive to the ICC in the Whole Health Team arm. In these two arms, an increased number of clinicians, more evenly distributed workload of clinicians or more homogeneous treatment group sizes leads to increased power. Our simulation study for the ICC estimation indicates that the posterior mean ICC estimator has less bias when the true ICCs are large (i.e., 0.10), but when the ICC is small (i.e., 0.01), the posterior median ICC estimator is less biased.

Conclusion: Knowledge of ICCs and the structure of clustering is critical to the design of individually randomized group treatment trials with partially nested clusters. We demonstrate that the ICC of the Whole Health Team arm can affect power in the wHOPE trial. A Bayesian approach provides a flexible procedure for estimating the ICCs under complex scenarios. More work is needed to educate the research community about the individually randomized group treatment design and encourage publication of ICCs to help inform future trial designs.

Keywords

Background/Aims

It is possible for participants in individually randomized trials to receive group-based treatments. For example, multiple participants can receive treatment from the same health professional, or they can be treated in a group setting along with other participants. Such trials are referred to as individually randomized group treatment trials.[1–4] The outcomes measured for participants treated by the same health professional or in the same group are often more similar than those measured across different health professionals or groups, leading to a positive intraclass correlation coefficient (ICC). Such within-cluster correlation needs to be accounted for appropriately during both the design and analysis. If not, the trial risks being underpowered or having an inflated type I error rate.

Concerns about potential clustering induced by group treatment arose after the initial design of the Whole Health Options in Pain Education trial (wHOPE), a three-arm trial comparing two active interventions (Whole Health Team and Primary Care Group Education) to and Usual Primary Care (control). The primary aim is to reduce pain interference; the secondary aim is to improve functioning and quality of life among Veterans with moderate to severe chronic pain.[5] Individuals randomized to the Whole Health Team arm attend one initial and four follow-up visits with an interdisciplinary co-located team consisting of a medical provider and at least one complementary integrative health provider (e.g., acupuncturist, Tai Chi instructor) to develop a personalized health plan to manage chronic pain using non-pharmacological strategies. They also receive eight individual coaching sessions delivered by primarily one Whole Health coach to reinforce the personalized health plan. Altogether, this arm has a total of 31 primary and backup clinicians (8 primary care providers, 13 integrative health providers, 10 Whole Health coaches) across 5 geographically diverse

enrollment sites (details in Supplemental Table 1). Individuals randomized to Primary Care Group Education participate in five weekly open group cognitive behavioral therapy sessions for chronic pain led by one of 11 psychologists, each with potentially varying group sizes (usually 4–10 participants) and composition of participants, creating a complex interaction network. Such a multiple-to-multiple relationship has been previously discussed under the framework of multiple membership models in partially nested designs.[6–9] A sample individual-clinician interaction schedule for the two treatment arms is in Supplemental Table 2. The Usual Primary Care arm consists of routine primary care for chronic pain in which participants are individually managed by their own primary care providers. Clustering is not considered for this arm. Previous studies have independently investigated grouped treatment and ungrouped control.[10–12] The impact of complex intraclass correlations at the clinician and session levels on sample size and power, such as in wHOPE, has not yet been fully investigated in the trial design literature.

To address this gap, we discuss a multiple membership model that accommodates three arms with different nesting structures, as in the wHOPE trial. We investigate how the study power is affected by the magnitude of ICCs as well as variations in mapping matrices characterizing the dynamic participant-session and participant-clinician networks. Estimating ICCs in the multiple membership model may be more complex than standard mixed effects models; therefore, we present a Bayesian ICC estimator and examine its operating characteristics via simulations to inform potential sample size re-estimation for wHOPE based on interim data.

Methods

The wHOPE Trial

wHOPE is a pragmatic, individually randomized clinical trial to address pain management in Veterans with moderate to severe chronic pain. Figure 1 presents the wHOPE study design; full details have been published elsewhere.[5] Briefly, participants are randomized to three arms, two active treatments (Whole Health Team and Primary Care Group Intervention)[13] and Usual Primary Care in an approximately 11:11:2 ratio, stratified by site. The primary outcome is change in pain interference score using the Brief Pain Inventory[14] from baseline to the final study outcome assessment at 12 months. The score is an average of seven items (scale 0–10) measuring the extent to which pain has interfered with various domains such as enjoyment of life and general activity.

wHOPE plans to enroll 341, 341, and 63 participants (accounting for attrition) with effective sample sizes of 275, 275 and 50 participants in the Whole Health Team, Primary Care Group Education, and Usual Primary Care arms, respectively. With an overall 5% type I error (2-sided) and 90% power, sample sizes were estimated for each pairwise comparison with Bonferroni weighted type I error: Whole Health Team vs. Primary Care Group Education (α=0.03); Whole Health Team vs. Usual Primary Care (α=0.01); and Primary Care Group Education vs. Usual Primary Care (α=0.01). The detectable standardized effect size is 0.30 for the Whole Health Team vs. Primary Care Group Education and 0.60 for the other two comparisons (assuming a standard deviation of 2.4 for the Brief Pain Inventory).

Multiple Membership Model and Power for wHOPE

Multiple membership models have been developed for trials with dynamic treatment groups.[4,6] We extend the multiple membership model to three arms, as in wHOPE. One key feature of the Whole Health Team is that the Whole Health coach works within the interdisciplinary clinical team consisting of a medical provider and integrative health provider. The team follows participants' progress toward these goals over at least 4 subsequent clinical visits. Whole Health coaches also provide participants 8 weekly coaching sessions focused on helping them achieve their pain care goals. Primary Care Group Education consists of five core 90-minute group sessions on pain self-management led by a psychologist. Groups are open (individual participants may start and finish the five-session series as they are enrolled, not necessarily in order) and participants are expected to finish at least four of five sessions within 3 months of enrollment. In addition to the core sessions, there is an orientation and a discharge session. Supplemental Table 1 contains numbers and roles of clinicians in the two treatment arms.

For simplicity, we omit the subscript for treatment arm and discuss the sub-model for each arm. The control arm has no grouped treatment and can be simply expressed as

$$Y_i = \mu_1 + \epsilon_i$$

where Y_i is a continuous outcome for ith individual; ϵ_i is the sampling error following $N(0, \sigma^2)$; and μ_1 is the grand mean.

Assume the Whole Health Team arm has a total of J clinicians. Let b_j be the effect of clinician j that follows a normal distribution with mean 0 and variance τ^2, where τ^2 indicates the variation in performance among clinicians. Each participant is expected to complete a total of S study visits that can be delivered by different clinicians. A weighted mapping matrix W_{IJ} can map the contribution of clinicians to participants' treatment, where each cell w_{ij} represents the proportion of the S sessions delivered to individual i by clinician j, and $\sum_{j=1}^{J} w_{ij} = 1$ for all i. We represent the outcome in the Whole Health Team arm using a two-level model with a weighted random effect corresponding to each clinician:

$$Y_i = \mu_2 + \sum_{j=1}^{J} w_{ij} b_j + \epsilon_i.$$

We model the outcome for each participant in the Primary Care Group Education arm by accounting for both the session- and clinician-level random effects. Consider a total of L sessions and let matrix V_{I*L} be a weighted mapping matrix with each element $v_{i\ell}$ representing the weight for participant i allocated to session ℓ, where $v_{i\ell}$ will be either 0 or $\frac{1}{S}$ when participants are expected to attend S sessions, and $\sum_{\ell=1}^{L} v_{i\ell} = 1$ for all i. Let d_ℓ denote the session-level random effect. Define $\mathbb{Q}_\ell \in \{1, ..., J\}$ as the index for the clinician delivering session ℓ. Then, the random effect $d_\ell \sim N\left(\sum_{j=1}^{J} c_j \mathbb{1}_{\{j = \mathbb{Q}_\ell\}}, \pi^2\right)$, where c_j is the clinician effect following $N(0, \phi^2)$; π^2 and ϕ^2 are the variance components for sessions and clinicians, respectively. Then, the model can be written as

$$Y_i = \mu_3 + \sum_{\ell = 1}^{L} v_{i\ell}\, d_\ell + \epsilon_i.$$

Given we are interested in pairwise comparisons of each treatment arm, closed-form power formulas can be derived given: numbers of participants enrolled in each arm; target effect size; and ICCs due to group treatment. The power is inversely related to the variance of the treatment effect, $Var(\hat{\delta}_{k,k'})$, where $\hat{\delta}_{k,k'}$ corresponds to the difference in the outcome means between arm k and k'. Given the nominal type 1 error of α and hypothesized effect size of $\Delta_{k,k'} = |\mu_k - \mu^{k'}|$, the power $1 - \beta$ of a two-sided test is,

$$1 - \beta = \Phi\left(\frac{\Delta_{k,k'}}{\sqrt{Var(\hat{\delta}_{k,k'})}} - z_{\frac{\alpha}{2}} \right) \tag{1}$$

In the Appendix, we show that

$$Var(\hat{\delta}_{2,1}) = \eta_2^2\left(\frac{1 - \rho_1}{n_1} + \frac{1}{n_2}\left(1 + \left(\mu_w + \frac{\sigma_w^2}{\mu_w} - 1 \right)\rho_1 \right) \right),$$

$$Var(\hat{\delta}_{3,1}) = \frac{\eta_3^2}{n_3}\left(1 + \left(\mu_v + \frac{\sigma_v^2}{\mu_v} - 1 \right)\rho_2 \right) + \frac{\sigma^2}{n_1},$$

$$Var(\hat{\delta}_{3,2}) = \frac{\eta_3^2}{n_3}\left(1 + \left(\mu_v + \frac{\sigma_v^2}{\mu_v} - 1 \right)\rho_2 \right) + \frac{\eta_2^2}{n_2}\left(1 + \left(\mu_w + \frac{\sigma_w^2}{\mu_w} - 1 \right)\rho_1 \right),$$

where n_k is the number of participant per arm, $\eta_2^2 = \tau^2 + \sigma^2$ and $\eta_3^2 = \phi^2 + \pi^2 + \sigma^2$ are the total variances in the two treatment groups, and, μ_w and σ_w^2 and μ_v and σ_v^2 are the mean and variance of w_j (vector of the column sum of W) and v_j (vector of the column sum of V), respectively. We assume these quantities are known for design purposes. Across all variance expressions, the ICC for clinicians in the Whole Health Team is $\rho_1 = \frac{\tau^2}{\sigma^2 + \tau^2}$, and the total ICC for both sessions and clinicians in the Primary Care Group Education arm is $\rho_2 = \frac{\pi^2 + \phi^2}{\sigma^2 + \pi^2 + \phi^2}$. A common range (i.e., 0 to 0.1)[15] can be used to assess the impact of ICC on power or sample size estimates.

Estimating Statistical Power for wHOPE

Using formula (1) and the derived variance expressions, we calculate the power for the three contrasts in wHOPE. We first assume that only the coaching sessions conducted by the Whole Health coaches are responsible for the treatment effect in the Whole Health Team

arm. We consider two cases for the allocation of participants to coaches, each leading to W matrices with different column sums denoted W_1 and W_2. Under W_1, each participant is treated by one primary Whole Health coach, whereas under W_2, each participant can be treated by a backup coach, if available. In this case, we assume that the primary coach delivers 7 out of the 8 sessions, and the backup coach delivers 1 session, per protocol. The column sums for W_1 and W_2 are displayed in Table 1. We consider five ICCs, $\rho_1 \in \{0, 0.01, 0.02, 0.05, 0.1\}$.

Under a more relaxed, and probably more realistic assumption in which clinicians other than the Whole Health coaches (Supplemental Table 1) contribute to better pain management of participants, we assume the treatment effect is related to the number of encounters between participants and all clinicians available at each site. In the actual treatment delivery, all clinicians, including Whole Health coaches, will have an additional 0–4 encounters (excluding the initial visit) with each participant, depending on the site availability and arrangement with participants. We consider four extreme cases for the mapping of participants to clinicians, denoted as W_3, W_4, W_5 and W_6. Detailed specifications for the weight of each clinicians in these conditions are provided in Table 1.

For the Primary Care Group Education arm, we expect each participant to complete 5 sessions. Based on the expected session size of 4–10 participants, we expect a total of 200 sessions with an average of 6–7 participants per session across five sites. We also consider the mapping of participants to sessions in two extreme scenarios that will lead to the same mean (1.38) but different variances of the column sums of V. Under V_1, each session has an equal number of participants, and the variance of the column sum is 0; under V_2, each session has either 1 or 10 participants, and the variance of the column sum is 0.74. We consider four ICCs, $\rho_2 \in \{0, 0.01, 0.1, 0.2\}$.

Bayesian approach to estimating the ICC

Prior work based on the multiple membership model has considered the restricted maximum likelihood approach to estimate the treatment effect and correlations, an approach that can be implemented using SAS *PROC MIXED* or the *xtmixed* function in *Stata*.[6,8] However, applying this frequentist approach could be challenging for more complex scenarios, such as the three-level nesting structure of the Primary Care Group Education arm or jointly estimating a multiple-arm model with diverse and complex clustering structures, as in wHOPE.

In contrast, a Bayesian approach can provide a flexible alternative to estimating ICCs. Through Markov Chain Monte Carlo, the Bayesian approach directly provides a posterior sample and credible intervals for regression parameters and ICCs and can be extended to accommodate complex clustering. In many cases, posterior distributions can be efficiently sampled via the Gibbs sampler, providing similar or even more stable estimates compared to the frequentist approach in terms of bias and efficiency.[8] To assist in the reporting of ICCs in the interim or final analysis of wHOPE, we derived Bayesian correlation estimators for the Whole Health Team and the Primary Care Group Education arms, assuming the models introduced earlier. Though not pursued here, our sampler (see derivation in Appendix)

also allows for covariate adjustment (i.e., site). In what follows, we examine the empirical performance of these estimators in the context of the wHOPE study.

Simulation design based on wHOPE

We designed a simulation study closely resembling wHOPE to examine the performance of the Bayesian estimators of the ICCs within the multiple membership model. We assumed that each treatment arm includes 275 participants. For the Whole Health Team arm, we consider six mapping matrices $W_a – W_f$ following the description of $W_1 – W_6$ for patient-clinician interactions with the following additional condition. When multiple clinicians of the same role are available at a site (primary and backup combined), each participant has a probability of 0.75 of being assigned to only one clinician and 0.25 to two clinicians. At sites where two or more primary clinicians are available, the primary clinicians were randomly chosen. No partial clustering is allowed when only one available clinician per role exists at a site. With these specifications, we randomly generated W and the outcomes under five ICCs: 0.01, 0.02, 0.05, 0.10, and 0.15.

For the Primary Care Group Education arm, we assumed 40 sessions at each site and an average 6–7 participants per session. We considered the two scenarios with even and uneven participant-to-session mapping matrices, V. Each cell of the mapping matrix is either 0 or 0.2, and the row sum of V is 1. Given the difficulty in generating random matrices under constraints of both column and row sums, we generated the two scenarios as follows. In the first scenario (V_a), participants were evenly distributed across sessions; at each site, each participant was randomly allocated to 5 of 40 sessions. In the second scenario (V_b), participants were unevenly distributed; each participant was randomly assigned to 4 sessions in the first 15 sessions and 1 of the other 25 sessions using sampling without replacement. We further considered two session-to-clinician mapping matrices U: primary providers delivered 90 percent (U_a) or 60 percent (U_b) of the 40 sessions at each site with the backup providers covering the rest. Note that each session is delivered by only one provider and there is no mixed membership in session-to-clinician mapping. One site (site D) has multiple backup providers and will split the workload evenly. For both the session- and clinicians-level clustering, we assumed an ICC of 0.01 or 0.10. For both arms, the treatment effect was assumed to be –1.5, which is approximately 0.3 standard deviations of the outcome.

We simulated 1000 trials for each parameter combination and fit the Bayesian estimators by running a chain with 5,000 iterations and 2000 burn-ins. Traceplots indicated adequate convergence. Thinning by 20 steps was used to reduce the autocorrelation between updates. We used conjugate priors for all model parameters: an inverse Gamma distribution of IG(0.001,0.001) for all variance parameters and normal N(0,1000) for the mean parameters. The induced prior for the ICC parameter is U shaped (shown in Figure 2 of Speigelhater[16]) In each scenario, we estimated the posterior mean and median and their biases for the ICC. All simulations were performed in R 4.0.1 and the code can be found at https://github.com/ttyale/wHOPE-CT. We note that other commercial software, such as $MLwIN$[17] can be used for Bayesian estimation of the multiple membership model.

Results

Power calculation results for wHOPE

Figure 2 presents the power curves for varying combinations of W_1, W_2, V_1 and V_2 for the three pairwise comparisons under the largest $\rho_2 = 0.20$. In general, we found that the power for the contrasts is sensitive to the ICC in the Whole Health Team arm, but not to the ICC in the Primary Care Group Education. The detailed results for all scenarios are presented in Supplemental Table S3. The Primary Care Group Education vs. Usual Primary Care contrast is robust to the change in ICC and the mapping of participants to sessions and can persistently obtain power close to 90% even when ρ_2 is 0.20. The Whole Health Team vs. Usual Primary Care contrast is more sensitive to the change in ICC but is still able to obtain close to 80% power with ρ_1 near 0.05. The contrast of Whole Health Team vs. Primary Care Group Education is very sensitive to ρ_1. When ρ_1 exceeds 0.02, the power falls below 80%. Within the basic scenarios (W_1-W_2; V_1-V_2), the largest power is obtained under $W_2 V_1$, when treatment delivery is more evenly distributed over the Whole Health coaches and treatment sessions in Primary Care Group Education have equal sizes.

Under more relaxed assumptions (scenarios W_3-W_6), we focus on the Whole Health Team-related contrasts since they are sensitive to change in the ICC. Full results are included in Supplemental Tables S4 and S5. We observe slightly improved power across all scenarios compared to scenarios W_1 and W_2. The rank of power follows $W_3 < W_4 < W_5 < W_6$, a pattern consistent with the decreasing order of the variance of column sums of W matrices. This result suggests power improvement is associated with increased participant-clinicians encounters for all different roles and increased shares of contribution by backup clinicians.

Empirical evaluation of the Bayesian intraclass correlation estimator

Table 2 summarizes the average posterior mean and median estimators of the ICC and their absolute bias for the Whole Health Team arm. In general, the posterior mean estimator is close to the true value when the true correlation parameter is large (i.e.,0.10). With a small true ICC (i.e., 0.01), the posterior median estimator is on average less biased compared to the posterior mean estimator. This is likely because the true posterior distribution of the correlation is skewed when the correlation is close to the boundary, in which case the median estimator outperforms the mean estimator. The posterior mean often has upward bias when the true intraclass correlation is small.

Table 3 summarizes the results for estimating the ICCs for the Primary Care Group Education arm. Both the posterior mean and median estimators have small bias when the true ICCs at both session and clinician levels are large. When the true ICCs are small, the posterior median estimator appears less biased compared to the posterior mean estimator. The Bayesian estimators seem to have larger upward bias for the session-level correlation parameter compared to the clinicians-level correlation parameter. We found that the biases for the session-level ICC are related to the relatively small participant-to-session ratio (1.25:1) and an additional simulation (Supplemental Table S6) with 2,500 participants (participant-to-session ratio is 12.5:1) found negligible bias for estimating the session-level ICC even for small ICCs. Finally, among the mapping matrices U and V we considered,

there were no systematic patterns describing how these variations affected the bias of the Bayesian ICC estimators.

Discussion

Motivated by the wHOPE trial, we studied how complex clustering induced during group treatment can affect study power. We demonstrated that the ICC in the Whole Health Team arm can have a significant impact on power especially for the comparison of Whole Health Team vs. Primary Care Group Education. An ICC of 0.02 or above could lead to less than 80% power for that contrast in various partial clustering scenarios. However, preliminary data from the VOICE trial (NCT: 03026790) that implemented an Integrated Pain Team treatment, an intervention very similar to the Whole Health Team intervention, yielded an ICC across sites (a proxy for clinicians) of 0.007, which suggests that wHOPE may very likely maintain over 80% power to detect the effect estimate in the aforementioned contrast. Nevertheless, close monitoring of the ICC in the Whole Health Team arm during data collection is recommended to allow for sample size re-estimation[18,19] should the power no longer be sufficient. Future work not only needs to focus on how best to monitor this ICC (i.e., methods, timing), but also on the role site variability plays in the estimation of the ICC. The studies we present in this paper did not separate site effects from clinician effects, but to truly understand the impact of the group treatments on the power of the study, we plan to conduct future methodological investigations to separate and properly account for these effects.

At the planning stage, an increased number of clinicians, especially for the Whole Health Team arm, will lead to larger power, but may be more expensive and may not be clinically in the best interest of patients who typically prefer to work with the same clinician teams. To plan an optimal design that attains the desired power within budget constraints or under the uncertainty of ICC,[20,21] knowledge of the mapping matrices is essential, especially in designs with partially clustered treatment groups. One important finding from wHOPE is that more evenly distributed workload across clinicians or more balanced treatment group sizes would lead to improved power. When the expected participant-to-clinician ratio, μ_w, and the expected participant-to-session ratio, μ_v, are known, our variance formulas suggest that maximum power will be obtained when $\sigma_w^2 = \sigma_v^2 = 0$. This corresponds to a scenario in which participants are evenly allocated across clinicians or sessions and there is minimum variability within each column of the mapping matrices. In general cases, where $\sigma_w^2, \sigma_v^2 > 0$, our empirical evaluations suggest that power for treatment comparisons is lower for larger values of σ_w^2 and σ_v^2. This suggests controlling for σ_w^2 and σ_v^2 in trial implementation is an alternative way to reduce variance inflation due to group treatment.

In addition to addressing the statistical power for treatment comparisons, we also examined the empirical performance of Bayesian ICC estimators within the multiple membership model to guide future reporting of these estimates in the wHOPE study. We demonstrated that we can reasonably estimate ICCs for the proposed models with two levels (e.g., Whole Health Team) and three levels (e.g., Primary Care Group Education), and found that the posterior median estimator may have lower bias than the posterior mean estimator for

estimating a small ICC parameter (as we expect in wHOPE). Similar patterns have also been observed in previous simulation studies,[6] and it is generally agreed that unbiased estimation is more difficult to achieve when the true value is close to the boundary. Our Bayesian estimators can be applied to other partially nested trials to obtain empirical ICC estimates to guide the planning of future studies with a similar endpoint (as exemplified in Table 11.1 of Moerbeek and Teerenstra[22] for cluster randomized trials). Furthermore, incorporating more informative priors on variance or ICC parameters based on existing studies[23,24,25] may lead to more precise posterior estimations in a partially nested design.

Finally, while our current evaluation assumes separate models for each arm, we plan to extend our approach to jointly estimate parameters in multiple arms with different levels of clustering and to a broader type of partially clustered model. For example, models with role-specific random effects could be postulated, such as by clinicians in different roles in the Whole Health Team arm, or between clinicians and sessions (e.g., Primary Care Group Education) when each session is co-delivered by multiple providers. For any of these models, covariate adjustment can be included. Overall, the flexibility of Bayesian estimation may lead to an advantage when complex, partially clustered structures need to be accounted for in a trial.

Conclusion

The wHOPE trial served as an instructive case example to demonstrate the impact of ICCs on the power of trials with individually randomized group treatment and partially nested clusters. Power is improved with an increased number of clinicians or when participants are more evenly allocated across clinicians or treatment sessions. We also investigated the reliability of a Bayesian approach to estimate the ICC efficiently through simulation studies. This case study represents an opportunity to increase awareness of individually randomized group treatment designs and to encourage publication of ICCs to help better guide future designs of these trials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement:

APPENDIX

Derivation of the Power Formulae

Power and sample size estimation are derived via the pairwise comparison approach. We estimate the contrast of each pair of treatment arms $\delta_{k,k'}$ through calculating means and variances in each arm. For simplicity, we assume the error variance, denoted σ^2, is homogeneous across all three arms. The Usual Primary Care, Whole Health Team, and Primary Care Group Education arm each has n_1, n_2 and n_3 participants with mean outcome $\bar{Y}_1$, $\bar{Y}_2$, and $\bar{Y}_3$, respectively. For the Usual Primary Care arm, the mean is $\bar{Y}_1 = \mu_1 + \frac{1}{n_1}\sum_{i=1}^{n_1}\epsilon_i$ and its variance is $Var(\bar{Y}_1) = \frac{\sigma^2}{n_1}$.

For the Whole Health Team arm, assume there are J_2 clinicians. The outcome mean is

$$\bar{Y}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2}\left(\mu_2 + \sum_{j=1}^{J_2} w_{ij}b_j\right) = \mu_2 + \frac{1}{n_2}\sum_{j=1}^{J_2}\left(b_j\sum_{i=1}^{n_2} w_{ij}\right) + \frac{1}{n_2}\sum_{i=1}^{n_2}\epsilon_i$$

Let $w_j = \sum_{i=1}^{n_2} w_{ij}$, and assume the w_j and clinicians-specific random effect are independent. Then the variance of $\bar{Y}_2$ is,

$$Var(\bar{Y}_2) = \frac{\tau^2\sum_{j=1}^{J_2} w_j^2 + n_2\sigma^2}{n_2^2}$$

To simplify this variance, if we assume the mean and variance of w_j are μ_w and σ_w^2, then $\mu_w = \frac{n_2}{J_2}$, and $\sum_{j=1}^{J_2} w_j^2 = J_2(\mu_w^2 + \sigma_w^2)$. Let the ICC across clinicians be $\rho_1 = \frac{\tau^2}{\tau^2 + \sigma^2}$, and the total variance be $\eta_2^2 = \tau^2 + \sigma^2$, then we have,

$$Var(\bar{Y}_2) = \frac{\eta_2^2}{n_2}\left(1 + \left(\mu_w + \frac{\sigma_w^2}{\mu_w} - 1\right)\rho_1\right)$$

The treatment effect of the Whole Health Team arm over the Usual Primary Care is then $\hat{\delta}_{2,1} = \bar{Y}_2 - \bar{Y}_1$, and the variance is,

$$Var(\hat{\delta}_{2.1}) = Var(\bar{Y}_2) + Var(\bar{Y}_1) = \eta_2^2\left(\frac{1 - \rho_1}{n_1} + \frac{1}{n_2}\left(1 + \left(\mu_w + \frac{\sigma_w^2}{\mu_w} - 1\right)\rho_1\right)\right)$$

For the type I error level α, the power $(1 - \beta)$ for detecting an effect of $\delta_{2,1}$, is given by,

$$1 - \beta = \Phi\left(\frac{\Delta_{2,1}}{\sqrt{Var(\hat{\delta}_{2,1})}} - z_{\alpha/2}\right)$$

Here, $\delta_{2,1}$ is the design effect. The corresponding sample size formula to reach the level of power $1 - \beta$ can also be easily obtained by inverting the above equation.

We now consider the outcome and variance estimation for the Primary Care Group Education arm with clustering due to sessions and clinicians. Assume there are L_3 sessions and J_3 clinicians. We can obtain the outcome mean,

$$\bar{Y}_3 = \mu_3 + \frac{1}{n_3} \sum_{\ell=1}^{L_3} \left(d_\ell \sum_{i=1}^{n_3} v_{i\ell}\right) + \frac{1}{n_3} \sum_{i=1}^{n_3} \epsilon_i$$

Recall that $d_\ell \sim N\left(\sum_{j=1}^{J} c_j \mathbb{1}\{j = \mathbb{Q}_\ell\}, \pi^2\right)$ denotes the session-level random effect for relevant therapist j, where c_j is the clinician effect following $N(0, \phi^2)$; $\mathbb{Q}_\ell \in \{1, ..., J\}$ is the index for the clinician delivery session ℓ Assume the patient-to-session mapping matrix is V with elements of $v_{i\ell}$ and let the column sum of V be $v_\ell = \sum_{i=1}^{n_3} v_{i\ell}$. Also, assume the mapping from sessions to clinicians is represented by matrix U, and its element $u_{\ell j}$ equals to either 1 or 0 indicating whether a session is led by a clinician. Given no partial clustering at the clinician level, we then know $\sum_{j=1}^{J_3} u_{\ell j}^2 = 1$. The variance of $\bar{Y}_3$ is,

$$Var(\bar{Y}_3) = \frac{\sum_{\ell}^{L_3} (v_\ell^2(\phi^2 \sum_{j=1}^{J_3} u_{\ell j}^2 + \pi^2)) + n_3\sigma^2}{n_3^2} = \frac{(\phi^2 + \pi^2)\sum_{\ell=1}^{L_3} v_\ell^2 + n_3\sigma^2}{n_3^2}.$$

Note that this expression would not hold if partial clustering exists at the clinician level. Partial clustering would lead to $\sum_{j=1}^{J_3} u_{\ell j}^2 \leq 1$ and reduction in the total variance. Again, to simplify $Var(\bar{Y}_3)$, assume the mean and variance of v_ℓ are μ_v and σ_v^2, and $\mu_v = \frac{n_3}{L_3}$. We can then have $\sum_{\ell}^{L_3} v_\ell^2 = L_3(\mu_v^2 + \sigma_v^2)$. Further, let the ICCs between clinicians be $\rho_2^A = \frac{\phi^2}{\phi^2 + \pi^2 + \sigma^2}$ and between sessions be $\rho_2^B = \frac{\pi^2}{\phi^2 + \pi^2 + \sigma^2}$. Let the total sum of variance be $\eta_3^2 = \phi^2 + \pi^2 + \sigma^2$. Then,

$$Var(\bar{Y}_3) = \frac{\eta_3^2}{n_3}\left(1 + \left(\mu_v + \frac{\sigma_v^2}{\mu_v} - 1\right)(\rho_2^A + \rho_2^B)\right)$$

We can now estimate the treatment effect of Primary Care Group Education over the Whole Health Team or the control arm. Define $\delta_{3,1} = \bar{Y}_3 - \bar{Y}_1$, and $\delta_{3,2} = \bar{Y}_3 - \bar{Y}_2$. The respective variances are then,

$$Var(\hat{\delta}_{3,1}) = \frac{\eta_3^2}{n_3}\left(1 + \left(\mu_v + \frac{\sigma_v^2}{\mu_v} - 1\right)(\rho_2^A + \rho_2^B)\right) + \frac{\sigma^2}{n_1}$$

$$Var(\hat{\delta}_{3,2}) = \frac{\eta_3^2}{n_3}\left(1 + \left(\mu_v + \frac{\sigma_v^2}{\mu_v} - 1\right)(\rho_2^A + \rho_2^B)\right) + \frac{\eta_2^2}{n_2}\left(1 + \left(\mu_w + \frac{\sigma_w^2}{\mu_w} - 1\right)\rho_1\right)$$

For the type I error rate of α, the formulae for detecting effect sizes $\Delta_{3,1}$ and $\Delta_{3,2}$ with power of $1 - \beta$ are given by,

$$1 - \beta = \Phi\left(\frac{\Delta_{3,1}}{\sqrt{Var(\hat{\delta}_{3,1})}} - z_{\alpha/2}\right)$$

$$1 - \beta = \Phi\left(\frac{\Delta_{3,2}}{\sqrt{Var(\hat{\delta}_{3,2})}} - z_{\alpha/2}\right)$$

Sample size formulae for given power can be inverted from the above equations.

Gibbs Sampler for Whole Health Team Arm

We provided the Markov Chain Monte Carlo sampler for the estimation of ICCs for the Whole Health Team. For simplicity, we re-express the outcome model for this arm in matrix form as

$$Y = X\beta + WB + E$$

Here, Y is the outcome vector, X is an n_2 by p design matrix including the intercept, and β is the vector of covariate effects. This general form of X allows for covariate adjustment such as the site effect. When there is no covariate, X is a vector of 1 and β is the overall mean. W is an n_2 by J_2 mapping matrix between participants and clinicians. B is a vector of length J_2 with $b_1, \cdots, b_{J_2}$ indicating clinician effects following a multivariate normal distribution, $N_{J_2}(0, \tau^2 I_{J_2})$. E is the residual vector following $N_{n_2}(0, \sigma^2 I_{n_2})$. Assume the prior distribution of β as $N_p(\beta_0, \sigma_0^2 I_p)$; the prior for σ^2 as inverse Gamma distribution $IG(a_0, b_0)$; the prior for τ^2 also as inverse Gamma $IG(c_0, d_0)$. The posteriors can be updated with the full conditionals as follows:

1. Sample B from multivariate normal posterior $N_{J_2}(M, V)$, where

$$M = \sigma^{-2}V W^T(Y - X\beta), \text{ and } V = \left(\sigma^{-2}W^T W + \tau^{-2}I\right)^{-1}$$

2. Sample β from normal posterior $N_p(M, V)$, where

$$M = V\left[\sigma^{-2}X^T(Y - WB) + \sigma_0^{-2}\beta_0\right] \text{ and } V = \left(\sigma^{-2}X^TX + \sigma_0^{-2}I\right)^{-1}$$

3.   Sample σ^2 from the posterior distribution,

$$IG(a_0 + \frac{n_2}{2}, b_0 + \frac{1}{2}(Y - X\beta - WB)^T(Y - X\beta - WB))$$

4.   Sample τ^2 from the posterior distribution,

$$IG(c_0 + \frac{J_2}{2}, d_0 + \frac{1}{2}B^TB)$$

With the posteriors, the ICC can be obtained by calculating $\dfrac{\tau^2}{\tau^2 + \sigma^2}$ for each iteration. A 95% credible interval for the ICC can also be obtained.

Gibbs Sampler for the Primary Care Group Education Arm

We provide the Markov Chain Monte Carlo sampler for the estimation of ICCs for the Primary Care Group Education arm. The model for this arm can also be expressed in the matrix form,

$$Y = X\beta + VD + E$$

Here, Y is the outcome vector, X is the 2 by p design matrix including the intercept, and β is the vector of regression coefficients. When there is no covariate, X is a vector of 1 and β is the overall mean. V is an n_3 by L_3 mapping matrix between participants and sessions. D is a vector of length L_3 with elements of $d_1, \cdots, d_{L_3}$ as session-specific effects. Here, sessions are nested within clinicians, and their mapping is denoted by U, a L_3 by J_3 matrix. Then, we have $D = UC$, where C is a vector of length J_3 for the clinician effect. C follows multivariate normal $N_{J_3}(0, \phi^2 I_{J_3})$. Then, D follows the multivariate normal distribution, $N_{L_3}(UC, \pi^2 I_{L_3})$. E is the residual vector follows $N_{n_3}(0, \sigma^2 I_{n_3})$. Assume the prior distribution of β is $N_p(\beta_0, \sigma_0^2 I_p)$; the prior for σ^2 follows an inverse Gamma distribution, $IG(a_0, b_0)$; the priors for π^2 for ϕ^2 also follow inverse Gamma, $IG(e_0, f_0)$ and $IG(g_0, h_0)$, respectively. The posteriors can be obtained with the full conditionals as follows:

1.   Sample D from multivariate normal $N_{L_3}(M*, V*)$, where

$$M* = V*\left[\sigma^{-2}V^T(Y - X\beta) + \pi^{-2}UC\right] \text{ and } V* = \left(\sigma^{-2}V^TV + \pi^{-2}I\right)^{-1}$$

2.   Sample β from multivariate normal $N_p(M*, V*)$, where

$$M* = V*\left[\sigma^{-2}X^T(Y - VD) + \sigma_0^{-2}\beta_0\right] \text{ and } V* = \left(\sigma^{-2}X^TX + \sigma_0^{-2}I\right)^{-1}$$

3. Sample C from the multivariate normal $N_{J_3}(M^*, V^*)$, where

$$M^* = \pi^{-2} V^* U^T D \text{ and } V^* = \left(\pi^{-2} U^T U + \phi^{-2} I\right)^{-1}$$

4. Sample σ^2 from the posterior distribution,

$$IG\left(a_0 + \frac{n_3}{2}, b0 + \frac{1}{2}(Y - X\beta - VD)^T(Y - X\beta - VD)\right)$$

5. Sample π^2 from the posterior distribution,

$$IG\left(e_0 + \frac{L_3}{2}, f_0 + \frac{1}{2}(D - UC)^T(D - UC)\right)$$

6. Sample ϕ^2 from the posterior distribution,

$$IG\left(g_0 + \frac{J_3}{2}, h_0 + \frac{1}{2}C^T C\right)$$

The ICC for sessions and clinicians can be obtained by calculating $\dfrac{\pi^2}{\pi^2 + \phi^2 + \sigma^2}$ and

$\dfrac{\phi^2}{\pi^2 + \phi^2 + \sigma^2}$ in each iteration of the sampler. 95% credible intervals for then can also be

obtained. Note that here we express the U matrix in a general form, and this Gibbs sampler can accommodate the scenarios where mixed membership between sessions and clinicians also exists (i.e., each treatment session is delivered by multiple clinicians).

References

1. Murray DM, Taljaard M, Turner EL, et al. Essential ingredients and innovations in the design and analysis of group-randomized trials. Annu Rev Public Health 2020; 41: 1–19. [PubMed: 31869281]

2. Pals SL, Murray DM, Alfano CM, et al. Individually randomized group treatment trials: a critical appraisal of frequently used design and analytic approaches. Am J Public Health 2008; 98(8): 1418–1424. [PubMed: 18556603]

3. Lee KJ and Thompson SG. Clustering by health professional in individually randomised trials. BMJ 2005; 330(7483): 142–144. [PubMed: 15649931]

4. Roberts C and Roberts SA. Design and analysis of clinical trials with clustering effects due to treatment. Clin Trials 2005; 2(2): 152–162. [PubMed: 16279137]

5. Seal KH, Becker WC, Murphy JL, et al. Whole Health Options and Pain Education (wHOPE): a pragmatic trial comparing whole health team versus primary care group education to promote non-pharmacological strategies to improve pain, functioning and quality of life in veterans--rationale, methods and implementation. Pain Med 2020: 21(Suppl 2): S91–S99. [PubMed: 33313734]

6. Roberts C and Walwyn R. Design and analysis of non-pharmacological treatment trials with multiple therapists per patient. Stat Med 2013; 32(1): 81–98. [PubMed: 22865729]

7. Walwyn R and Roberts C. Therapist variation within randomised trials of psychotherapy: implications for precision, internal and external validity. Stat Methods Med Res 2010; 19(3): 291–315. [PubMed: 19608603]

8. Browne WJ, Goldstein H and Rasbash J. Multiple membership multiple classification (MMMC) models. Stat Model 2001; 1(2): 103–124.

9. Sterba SK. Partially nested designs in psychotherapy trials: A review of modeling developments. Psychother Res 2017; 27(4): 425–436. [PubMed: 26686878]

10. Moerbeek M and Wong WK. Sample size formulae for trials comparing group and individual treatments in a multilevel model. Stat Med 2008; 27(15): 2850–2864. [PubMed: 17960589]

11. Bauer DJ, Sterba SK and Hallfors DD. Evaluating group-based interventions when control participants are ungrouped. Multivar Behav Res 2008; 43(2): 210–236.

12. Esserman D, Zhao Y, Tang Y, et al. Sample size estimation in educational intervention trials with subgroup heterogeneity in only one arm. Stat Med 2013; 32(12): 2140–2154. [PubMed: 23172724]

13. Seal KH, Borsari B, Tighe J, et al. Optimizing pain treatment interventions (OPTI): A pilot randomized controlled trial of collaborative care to improve chronic pain management and opioid safety—Rationale, methods, and lessons learned. Contemp Clin Trials 2019; 77: 76–85. [PubMed: 30572163]

14. Tan G, Jensen MP, Thornby JI, et al. Validation of the Brief Pain Inventory for chronic nonmalignant pain. J Pain 2004; 5(2): 133–137. [PubMed: 15042521]

15. Eldridge SM, Ashby D, Feder GS, et al. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. Clin Trials 2004; 1(1): 80–90. [PubMed: 16281464]

16. Spiegelhalter DJ. Bayesian methods for cluster randomized trials with continuous responses. Stat Med 2001; 20(3): 435–452. [PubMed: 11180312]

17. Rasbash J, Browne W and Goldstein H. A user's guide to MLwiN v2.33. Published online 2014.

18. Lake S, Kammann E, Klar N, et al. Sample size re-estimation in cluster randomization trials. Stat Med 2002; 21(10): 1337–1350. [PubMed: 12185888]

19. van Schie S and Moerbeek M. Re-estimating sample size in cluster randomised trials with active recruitment within clusters. Stat Med 2014; 33(19): 3253–3268. [PubMed: 24719285]

20. Innocenti F, Candel MJ, Tan FE, et al. Optimal two-stage sampling for mean estimation in multilevel populations when cluster size is informative. Stat Methods Med Res 2021; 30(2): 357–375. [PubMed: 32940135]

21. Korendijk EJ, Moerbeek M and Maas CJ. The robustness of designs for trials with nested data against incorrect initial intracluster correlation coefficient estimates. J Educ Behav Stat 2010; 35(5): 566–585.

22. Moerbeek M and Teerenstra S. Power analysis of trials with multilevel data. CRC Press; 2015.

23. Baldwin SA and Fellingham GW. Bayesian methods for the analysis of small sample multilevel data with a complex variance structure. Psychol Methods 2013;18(2):151–164. [PubMed: 23148476]

24. Turner RM, Omar RZ and Thompson SG. Constructing intervals for the intracluster correlation coefficient using Bayesian modelling, and application in cluster randomized trials. Stat Med 2006; 25(9): 1443–1456. [PubMed: 16220510]

25. Turner RM, Thompson SG and Spiegelhalter DJ. Prior distributions for the intracluster correlation coefficient, based on multiple previous estimates, and their application in cluster randomized trials. Clin Trials 2005; 2(2): 108–118. [PubMed: 16279132]
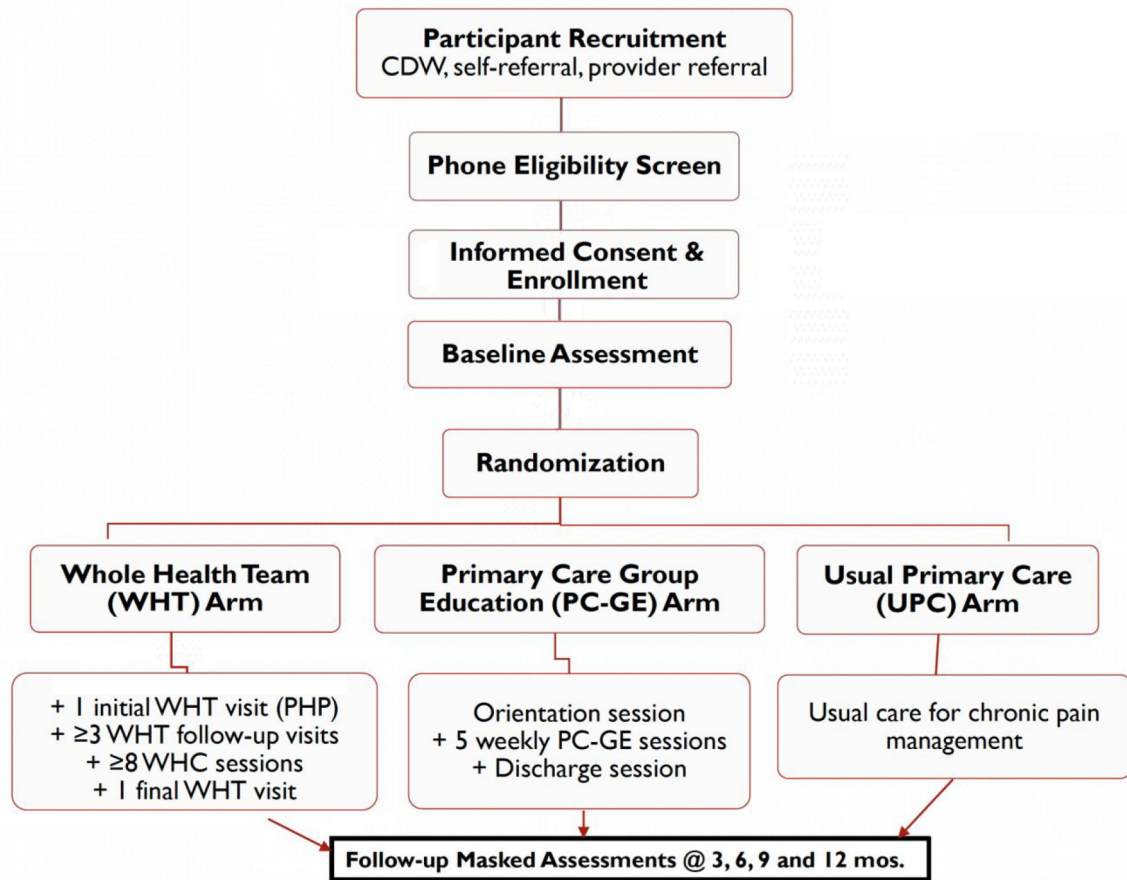
Figure 1.
Whole Health Team vs. Primary Care Group Education to Promote Non-Pharmacological Strategies to Improve Pain, Functioning, and Quality of Life in Veterans (wHOPE) Trial Flowchart.
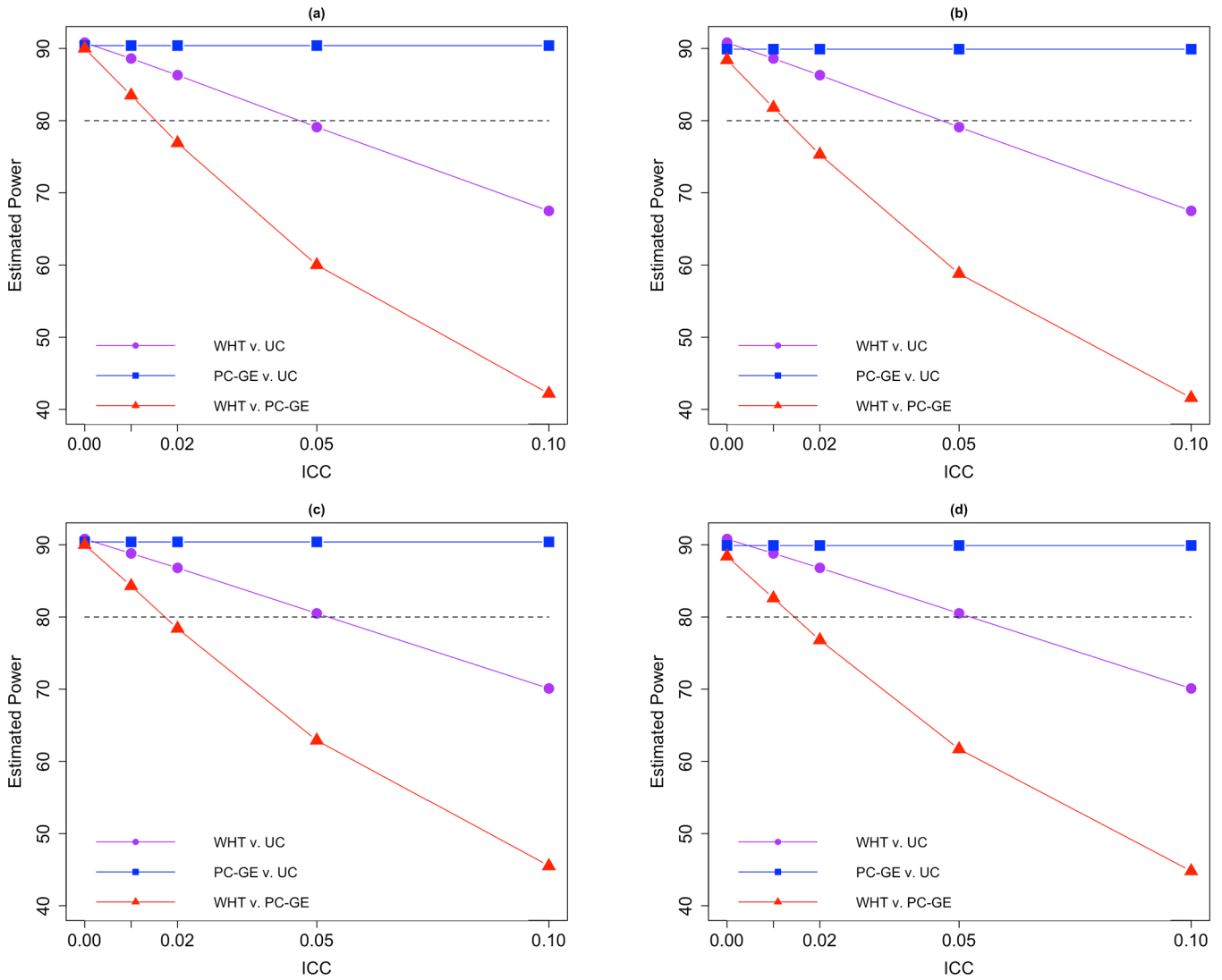
Figure 2.
Power curves for the three pairwise comparisons in the wHOPE trial (Whole Health Team vs. Usual Primary Care; Primary Care Group Education vs. Usual Primary Care; Whole Health Team vs. Primary Care Group Education) under varying ICCs in the Whole Health Team Arm and fixed ICC of 0.20 in Primary Care Group Education Arm.

WHT: Whole Health Team; PC-GE: Primary Care Group Education; UPC: Usual Primary Care (a): $W_1 V_1$ (b): $W_1 V_2$ (c): $W_2 V_1$ (d): $W_2 V_2$.

W_1: each participant is treated by only one primary Whole Health coach (8 sessions).

W_2: each participant is also treated by a backup coach (1 of 8 sessions) if backup is available.

V_1: each session has an equal number of participants.

V_2: each session has either 1 or 10 participants.

Table 1.

Column sums of the mapping matrix between participants and clinicians for each Whole Health coach and other clinicians under six different conditions.

W	Arm & Role	Site A	Site B	Site C	Site D	Site E
W_1	Primary WHC	55	55	55/3	55/2	55
	Backup WHC	0	--	--	--	0
W_2	Primary WHC	55*7/8	55	55/3	55/2	55*7/8
	Backup WHC	55*1/8	--	--	--	55*1/8
W_3	Primary PCP	55*4/15	55*4/14	55*4/15	55/2*4/15	55*4/15
	Backup PCP	0	0	--	--	--
	Primary CIH	55*1/15	55/2*1/14	55*1/15	55*1/15	55*1/15
	Backup CIH	0	0	--	--	--
	MH provider	55*1/15	--	55*1/15	55*1/15	55*1/15
	Primary WHC	55*9/15	55*9/14	55/3*9/15	55/2*9/15	55*9/15
	Backup WHC	0	--	--	--	0
W_4	Primary PCP	55*4/24	55*4/20	55*4/24	55/2*4/24	55*4/24
	Backup PCP	0	0	--	--	--
	Primary CIH	55*4/24	55/2*4/20	55*4/24	55*4/24	55*4/24
	Backup CIH	0	0	--	--	--
	MH provider	55*4/24	--	55*4/24	55*4/24	55*4/24
	Primary WHC	55*12/24	55*12/20	55/3*12/24	55/2*12/24	55*12/24
	Backup WHC	0	--	--	--	0
W_5	Primary PCP	55*4/15*3/4	55*4/14*3/4	55*4/15	55/2*4/15	55*4/15
	Backup PCP	55*4/15*1/4	55*4/14*1/4	--	--	--
	Primary CIH	55*1/15*3/4	55/2*1/14*3/4	55*1/15	55*1/15	55*1/15
	Backup CIH	55*1/15*1/4	55/2*1/14*1/4	--	--	--
	MH provider	55*1/15	--	55*1/15	55*1/15	55*1/15
	Primary WHC	55*9/15*31/36	55*9/14	55/3*9/15	55/2*9/15	55*9/15*31/36
	Backup WHC	55*9/15*5/36	--	--	--	55*9/15*5/36
W_6	Primary PCP	55*4/24*3/4	55*4/20*3/4	55*4/24	55/2*4/24	55*4/24
	Backup PCP	55*4/24*1/4	55*4/20*1/4	--	--	--
	Primary CIH	55*4/24*3/4	55/2*4/20*3/4	55*4/24	55*4/24	55*4/24
	Backup CIH	55*4/24*1/4	55/2*4/20*1/4	--	--	--
	MH provider	55*4/24	--	55*4/24	55*4/24	55*4/24
	Primary WHC	55*12/24*10/12	55*12/20	55/3*12/24	55/2*12/24	55*12/24*10/12
	Backup WHC	55*12/24*2/12	--	--	--	55*12/24*2/12

WHC: Whole Health Coach; PCP: primary care provider; CIH: complementary and integrative health; MH: mental health.

W_1: each participant would be treated by only one primary Whole Health coach (8 sessions).

W_2: each participant would also be treated by a backup coach (1 of 8 sessions) if such an arrangement is available at a specific site.

W_3: treatment is delivered by primary clinicians of all roles but no backup clinicians. Each participant maintains same team, with limited encounters (i.e., 4 times with primary provider, 1 time with all the other available roles including Whole Health coaches).

W_4: treatment is delivered by primary clinicians of all roles but no backup clinicians. The encounters for all roles are maximized (i.e., 4 times with primary provider and all the other roles if available).

W_5: treatment is delivered by both primary and backup clinicians of all roles. Other than the coaching sessions, the encounters are limited and each participant only meets clinicians of all different roles once. There is a 25% chance the encounters are not with the primary clinicians.

W_6: treatment is delivered by both primary and backup clinicians of all roles, with maximum encounters. All backup clinicians contribute to 1 out of the 4 sessions when they are available. For W_5 and W_6, we assume that 7 out of the 8 coaching sessions are delivered by the primary Whole Health coaches. Values are calculated based on 55 patients per site. "--" indicates no available clinician of a certain role at a site. "0" indicates a backup clinician of a certain role is not involved in the treatment. The number in each cell starts from the total number of patients, which is 55. The situation of 55/2 or 55/3 indicates two or three clinicians for a role exists at a site. The second term in each cell, if exists, is the weight based on the frequency of encounters for each role. The third term, if exists, further allocates weights between primary and backup clinicians within a role at each site.

Table 2.

Posterior mean and median of the ICC under different conditions in the Whole Health Team arm.

Clinicians Mapping	ICC	Pst. Mean	Mean Bias	Pst. Median	Median Bias
W_a	0.010	0.016	0.006	0.008	−0.002
	0.020	0.024	0.004	0.014	−0.006
	0.050	0.057	0.007	0.043	−0.007
	0.100	0.106	0.006	0.090	−0.010
	0.150	0.159	0.009	0.140	−0.010
W_b	0.010	0.016	0.006	0.008	−0.002
	0.020	0.023	0.003	0.014	−0.006
	0.050	0.052	0.002	0.039	−0.011
	0.100	0.106	0.006	0.089	−0.011
	0.150	0.162	0.012	0.144	−0.006
W_c	0.010	0.023	0.013	0.009	−0.001
	0.020	0.029	0.009	0.014	−0.006
	0.050	0.049	−0.001	0.031	−0.019
	0.100	0.093	−0.007	0.072	−0.028
	0.150	0.144	−0.006	0.122	−0.028
W_d	0.010	0.028	0.018	0.011	0.001
	0.020	0.032	0.012	0.013	−0.007
	0.050	0.049	−0.001	0.026	−0.024
	0.100	0.087	−0.013	0.062	−0.038
	0.150	0.129	−0.021	0.105	−0.045
W_e	0.010	0.025	0.015	0.010	0.000
	0.020	0.028	0.008	0.013	−0.007
	0.050	0.049	−0.001	0.030	−0.020
	0.100	0.087	−0.013	0.066	−0.034
	0.150	0.135	−0.015	0.113	−0.037
W_f	0.010	0.029	0.019	0.011	0.001
	0.020	0.032	0.012	0.014	−0.006
	0.050	0.050	0.000	0.029	−0.021
	0.100	0.084	−0.016	0.060	−0.040
	0.150	0.129	−0.021	0.105	−0.045

W_a-W_f are mapping conditions for clinicians in parallel to W_1-W_6 for the column sums of W in power calculation (see footnote in Table 1) with the following additional condition: each patient has 75% of the chance to only meet one clinician of each role when more than one clinicians of a certain role (primary and backup combined) is available at a site.

Table 3.

Posterior mean and median of the ICC under different conditions in the Primary Care Group Education arm.

Session/ Clinicians Mapping	Session ICC	Clinicians ICC	Session ICC Posterior				Clinicians ICC Posterior			
			Posterior Mean	Relative Bias for Posterior Mean	Posterior Median	Relative Bias for Posterior Median	Posterior Mean	Relative Bias for Posterior Mean	Posterior Median	Relative Bias for Posterior Median
V_aU_a	0.010	0.010	0.085	0.075	0.053	0.043	0.022	0.012	0.011	0.001
		0.100	0.102	0.092	0.068	0.058	0.117	0.017	0.097	-0.003
	0.100	0.010	0.093	-0.007	0.058	-0.042	0.026	0.016	0.014	0.004
		0.100	0.111	0.011	0.076	-0.024	0.128	0.028	0.107	0.007
V_aU_b	0.010	0.010	0.079	0.069	0.047	0.037	0.023	0.013	0.012	0.002
		0.100	0.100	0.090	0.065	0.055	0.094	-0.006	0.078	-0.022
	0.100	0.010	0.094	-0.006	0.060	-0.040	0.026	0.016	0.014	0.004
		0.100	0.117	0.017	0.082	-0.018	0.101	0.001	0.085	-0.015
V_bU_a	0.010	0.010	0.068	0.058	0.039	0.029	0.025	0.015	0.014	0.004
		0.100	0.090	0.080	0.058	0.048	0.118	0.018	0.097	-0.003
	0.100	0.010	0.083	-0.017	0.052	-0.048	0.030	0.020	0.018	0.008
		0.100	0.108	0.008	0.075	-0.025	0.127	0.027	0.106	0.006
V_bU_b	0.010	0.010	0.064	0.054	0.035	0.025	0.026	0.016	0.015	0.005
		0.100	0.090	0.080	0.058	0.048	0.094	-0.006	0.077	-0.023
	0.100	0.010	0.080	-0.020	0.049	-0.051	0.030	0.020	0.018	0.008
		0.100	0.110	0.010	0.078	-0.022	0.105	0.005	0.087	-0.013

V_a: participants are evenly distributed across sessions; at each site, each participant is randomly allocated to 5 of 40 sessions.

V_b: participants are unevenly distributed; each participant is randomly assigned to 4 sessions in the first 15 sessions and 1 of the other 25 sessions.

U_a: primary providers deliver 90 percent of 40 sessions at each site, and the backup providers cover the rest sessions.

U_b: primary providers deliver 60 percent of 40 sessions at each site, and the backup providers cover the rest sessions.