American Journal of Epidemiology
© The Author 2015. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.
Vol. 182, No. 12
DOI: 10.1093/aje/kwv132
Advance Access publication:
December 1, 2015

Practice of Epidemiology

Balancing Contamination and Referral Bias in a Randomized Clinical Trial: An Application of Pseudo-Cluster Randomization

Brian W. Pence*, Bradley N. Gaynes, Nathan M. Thielman, Amy Heine, Michael J. Mugavero, Elizabeth L. Turner, and Evelyn B. Quinlivan

* Correspondence to Dr. Brian W. Pence, Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, CB 7435, Chapel Hill, NC 27599-7435 (e-mail: bpence@unc.edu).

Initially submitted August 11, 2014; accepted for publication May 11, 2015.

In randomized trials of provider-focused clinical interventions, treatment allocation often cannot be blinded to participants, study staff, or providers. The choice of unit of randomization (patient, provider, or clinic) entails tradeoffs in cost, power, and bias. Provider- or clinic-level randomization can minimize contamination, but it incurs the equally problematic potential for referral bias; that is, because arm assignment of future participants generally cannot be concealed, differences between arms may arise in the types of patients enrolled. Pseudo-cluster randomization is a novel study design that balances these competing validity threats. Providers are randomly assigned to an imbalanced proportion of intervention-arm participants (e.g., 80% or 20%). Providers can be masked to the imbalance, avoiding referral bias. Contamination is reduced because only a minority of control-arm participants are treated by majority-intervention providers. Pseudo-cluster randomization was implemented in a randomized trial of a decision support intervention to manage depression among patients receiving human immunodeficiency virus care in the southern United States in 2010–2014. The design appears successful in avoiding referral bias (participants were comparable between arms on important characteristics) and contamination (key depression treatment indicators were comparable between usual care participants managed by majority-intervention and majority-usual care providers and were markedly different compared with intervention participants).

clinical trials; contamination; pseudo-cluster randomization; referral bias; study design

Abbreviations: HIV, human immunodeficiency virus; SLAM DUNC, Strategies to Link Antiretroviral and Antidepressant Management at Duke, University of Alabama-Birmingham, Northern Outreach Clinic, and the University of North Carolina at Chapel Hill.

For the highly valued randomized controlled trial study design, the validity of conclusions is dependent on at least 2 critical conditions. First, randomization must be implemented impartially, especially independent of knowledge of treatment arm, so that selection bias does not undermine the comparability of the treatment arms. Second, contamination must be minimized, so that participants actually receive the condition to which they were assigned and, in particular, those in the comparison arm do not receive the active intervention.

In considering these 2 conditions, researchers frequently face competing concerns in selecting the optimal unit of randomization in trials of health service delivery interventions. Although medication trials can use double-blinded, placebo-controlled designs to ensure that recruitment and patient and provider behavior are unaffected by knowledge of treatment arm, such blinding is often not possible for service delivery interventions. For example, if a trial focused on improving depression seeks to compare usual care with care by a depression "care manager" who guides the provider in prescribing and dosing antidepressants, neither the participant, provider, nor depression care manager can be blinded to treatment arm. If such a study were to implement patient-level randomization, with providers caring for both intervention and usual care participants, contamination would be a concern: Providers might apply principles from the depression care manager to improve their treatment of usual care participants. This contamination would tend to improve outcomes in the usual care arm, reducing the observed effect size and the trial's power

Downloaded from https://academic.oup.com/aje/article/182/12/1039/2195409 by guest on 25 April 2024

(1, 2). To avoid contamination, the trial might elect provider- or clinic-level (cluster) randomization, so that all patients of a given provider or clinic would be in 1 arm.

While addressing contamination, cluster randomization may incur a competing threat of selection or referral bias, because the treatment arm assignment of *future* enrollees cannot be masked from providers or staff (at least after the assignment of that provider's first participant is revealed). Unless a study's accrual can be completed before any treatment arm assignments are made, treatment arm will already be known for most participants before they are approached for enrollment. This foreknowledge may guide providers or staff, consciously or subconsciously, to recruit with different vigor or to enroll different types of patients in the arms, erasing the critical condition of comparability between arms on which causal inference depends (3, 4). For example, a cluster-randomized trial of a comparison of a new palliative care intervention with conventional care found, after the fact, that different types of patients had been enrolled into the 2 arms because of the clinical staff's awareness of arm assignment (5). Similarly, systematic reviews of hip protector trials have found that individually randomized trials have had mixed results, while cluster-randomized trials have consistently shown substantial benefits of hip protectors, likely due to preferential selection of patients into the intervention arms who were most likely to benefit (6, 7). Cluster randomization is also less efficient (requires a larger sample size) than individual randomization because of the design effect or the dependence of outcomes for participants within a cluster (1, 8).

Pseudo-cluster randomization is a novel study design developed to balance these competing threats of contamination and referral bias (9). Originally developed for the Dutch EASYCare Study, a trial of a new primary care management strategy for older adults compared with usual care, the design uses random assignment of providers to an imbalanced case mix (e.g., 80% intervention/20% usual care or 20% intervention/80% usual care). Contamination is reduced, although not eliminated, because most usual care arm participants are cared for by a provider with minimal exposure to the intervention. Referral bias is reduced because all participants have a nonzero probability of assignment to either arm and can be further avoided by masking providers and study staff to the use of imbalanced randomization. Pseudo-cluster randomization also improves efficiency relative to classic cluster randomization. Here, we describe the rationale for and implementation of pseudo-cluster randomization in a randomized trial of a depression care management strategy, compared with usual care, among patients with depression engaged in human immunodeficiency virus (HIV) care, known as the Strategies to Link Antiretroviral and Antidepressant Management at Duke, University of Alabama-Birmingham, Northern Outreach Clinic, and the University of North Carolina at Chapel Hill (SLAM DUNC) Study.

METHODS

The SLAM DUNC Study

This study is a 4-site randomized trial of the effect of depression treatment on antiretroviral medication adherence among HIV-infected patients with major depression (10). Participants in the intervention arm received Measurement-Based Care. Measurement-Based Care involves management by a depression care manager who uses regular assessments of key depression treatment metrics (depressive severity, medication adherence, medication side effects) and an evidence-based decision algorithm to recommend treatment changes to the treating HIV medical provider (11). For participants in the usual care arm, the treating HIV medical provider was free to refer or implement any depression treatment plan but did not receive support from the depression care manager.

In designing the study, we recognized limitations with both patient-level and provider-level randomization. Patient-level randomization raised the potential for contamination as described earlier. For this reason, the study was originally designed and funded with a provider-randomized design. However, in the protocol development phase, the study team became concerned about the potential for referral bias. This potential would arise because randomization could not be delayed until study accrual was complete, given the expected pace of recruitment and the desire not to delay initiation of treatment of the presenting depressive episode, and providers would therefore be aware of their arm assignment during recruitment. Providers were expected to have a strong preference for the intervention over usual care. As a result, with funder approval, the study team changed the design to pseudo-cluster randomization.

Pseudo-cluster randomization

In the Dutch EASYCare Study, Teerenstra et al. (9) developed a novel modification of the traditional provider-randomized design, called pseudo-cluster randomization, to balance contamination and referral bias. In pseudo-cluster randomization, each provider is randomly assigned not to an arm but to a specific majority/minority case mix. In the Dutch EASYCare Study, for example, each provider was randomly assigned to either an 80% intervention/20% usual care case mix or a 20% intervention/80% usual care case mix. Each presenting patient was randomly assigned upon recruitment, with assignment probabilities defined by the provider's case mix assignment. The potential for referral bias is reduced because each patient has a nonzero probability of receiving the intervention. Further, the case mix proportions—or even the use of imbalanced proportions—can be masked, and if the number of patients per provider is small, it will be difficult for the provider to suspect imbalanced randomization or to guess their assignment. Yet the design also addresses contamination. Usual care patients of majority-intervention arm providers may receive some components of the intervention condition, but these patients represent only a small fraction of all usual care patients. Usual care patients of majority-usual care arm providers are unlikely to receive intervention components because the exposure of the provider to the intervention is minimal.

Pseudo-cluster randomized designs can be analyzed by using the same statistical techniques as cluster-randomized designs, although alternatives have also been proposed (12). Efficiency is improved relative to a classic cluster-randomized design because the design effect is reduced (9). For a classic

Downloaded from https://academic.oup.com/aje/article/182/12/1039/2195409 by guest on 25 April 2024

cluster-randomized design, sample size is the product of the required sample size with an individually randomized study multiplied by the design effect, $D = 1 + (n - 1)\rho$ (8), where n is the mean cluster size and ρ is the intracluster correlation coefficient. In a pseudo-cluster randomized study, the design effect, D_{pc}, depends additionally on f, the case-mix proportion within clusters: $D_{pc} = [1 + (n - 1)\rho]/[1 + 4f(1 - f)n\rho/(1 - \rho)]$. In most cases, D_{pc} is substantially smaller than D (9). These formulas assume balanced cluster sizes; in the presence of unbalanced cluster sizes, power is likely to be reduced (13).

Sample size determination

The SLAM DUNC Study power calculations were designed to detect a 10 percentage point improvement in antiretroviral medication adherence, measured as a continuous variable, at 6 months after randomization. Previous studies suggested that the outcome would have a pooled standard deviation of between 25% and 29% (14), implying a Cohen's effect size of 0.345–0.400. For individual-level randomization, with 2-tailed 5% type I error probability and 80% power, the required sample size was 268 (134 per arm) for the smaller (conservative) effect size of 0.345. For classic provider-level randomization, the design effect was calculated as $D = 1.16$, based on estimates of 9 patients per provider and $\rho = 0.02$ (15–18). Thus, the required combined sample size for a classic provider randomized design would be $268 \times 1.16 = 312$. This number, increased to 390 in anticipation of 20% loss to follow-up at 6 months, was the originally funded sample size.

In implementing pseudo-cluster randomization, the study team selected a case-mix proportion (f) of 0.8, selected to balance contamination against the risk of unmasking the unbalanced design to providers and staff. This choice reduced the design effect to $D_{pc} = 1.04$. However, it was recognized that some contamination might occur, even though contamination would be reduced relative to patient-level randomization. Given that contamination would require substantial effort on providers' part (to replicate the depression care manager's schedule of regular systematic assessments) and that only 20% of usual care participants would be exposed to majority-intervention arm providers, the study team anticipated that a reduction in the effect size of 5%–10% due to contamination was realistic. This implied an effect size of 0.310–0.380 depending on the amount of contamination (5%–10%) and pooled standard deviation (25%–29%). The originally funded sample of 312 (corresponding to 300 in an individual design multiplied by $D_{pc} = 1.04$) therefore provided ≥80% power for most scenarios (effect size ≥ 0.328) and 76% power for the worst-case scenario (effect size = 0.310, corresponding to 10% contamination and pooled standard deviation of 29%). Given funding constraints, the originally funded sample size was therefore maintained.

Randomization

Providers were randomized 1:1 to either a majority (80%) intervention or majority usual care case mix. Randomization was stratified by site (3 levels; the smallest site was combined with another site because of shared providers) and by experience and comfort with antidepressant management (divided into high/moderate/low tertiles), as it was anticipated that provider experience and comfort would affect treatment quality and therefore participant treatment response, especially in the usual care arm. Experience and comfort with antidepressant management were assessed via a semistructured baseline provider interview assessing providers' standard practice in identifying depression and initiating and managing antidepressant treatment (19). Provider randomization was implemented in blocks of 4 within each site/experience stratum. For the purposes of randomization, Infectious Diseases fellows who managed patients in consultation with an attending physician were considered 1 unit with their attending physician.

Once providers were assigned to case mixes, randomization sequences for each provider's patients were determined in blocks of 10. For example, a majority intervention provider would have a sequence with 8 intervention arm slots and 2 usual care arm slots, in random order. Randomization sequences were generated and maintained by the study statistician and were available only to limited study staff who were not involved in recruitment activities. When a new participant enrolled in the study, recruitment staff would call the coordinating center, report certain information including the participant's provider, and would be given the participant's arm assignment. At 1 site, each patient was managed by a team of an attending physician and a mid-level provider or fellow; participants at this site were randomized on the basis of the case mix of their team's attending physician. Regular quality assurance was conducted to ensure that slots were correctly assigned in the specified order.

Blinding and ethical considerations

In the institutional review board protocol, the study team indicated the intention to use pseudo-cluster randomization but to blind participants, providers, and nearly all study staff to the pseudo-cluster aspect of the design. Blinding to the design was perceived as necessary to avoid referral bias. Blinding was further expected to pose negligible risk of harm to participants or providers because all enrolling participants would have a nonzero probability of being assigned to either arm. The principal investigator and co-principal investigator, a small number of senior clinical investigators, and study staff handling institutional review board correspondence were the only ones informed of the pseudo-cluster aspect of the design. Patient and provider consent forms both stated that patients would be "randomly assigned" to 1 of 2 arms. The study team planned to disclose the pseudo-cluster design to participating providers after the conclusion of the study.

Analysis

The study's Statistical Analysis Plan prespecified that analyses would be intention to treat and would address the pseudo-cluster randomized design. The intervention effect is estimated as the difference between arms in the primary (continuous) outcome at 6 months, estimated from a linear mixed model fit to a data set with one 6-month observation per participant, including an indicator variable for study arm, fixed effects for stratification variables (site and provider

Downloaded from https://academic.oup.com/aje/article/182/12/1039/2195409 by guest on 25 April 2024

Table 1. Characteristics of HIV Providers, SLAM DUNC Study, Southern United States, 2010–2014

Characteristic	Majority Intervention Providers ($n = 28$)			Majority Usual Care Providers ($n = 25$)		
	No.	%	Mean (SD)	No.	%	Mean (SD)
Stratification factors						
Site						
1 and 4	11	39		9	36	
2	9	32		8	32	
3	8	29		8	32	
Depression treatment practices score (range, 0–10)			5.3 (2.1)	5.5	2.5	5.5 (2.5)
High	9	32		10	40	
Medium	9	32		8	32	
Low	10	36		7	28	
Other						
Training						
Attending physician	23	82		21	84	
Mid-level provider[a]	5	18		4	16	
Clinical effort, %			38 (26)			37 (20)
% of clinical effort devoted to HIV			64 (34)			71 (28)
Years treating HIV			12.9 (7.4)			14.5 (9.4)
Confidence prescribing antidepressants (range, 1–5)[b]			3.2 (1.0)			3.7 (0.8)
Confidence changing antidepressants (range, 1–5)[b]			2.4 (0.8)			2.9 (0.9)
Believes treating depression is part of role (range, 1–5)[b]			3.9 (1.0)			3.9 (1.0)

Abbreviations: HIV, human immunodeficiency virus; SD, standard deviation; SLAM DUNC, Strategies to Link Antidepressant and Antiretroviral Management at Duke, University of Alabama-Birmingham, Northern Outreach Clinic, and the University of North Carolina at Chapel Hill.

[a] Nurse practitioner or physician assistant.

[b] On a1–5 scale, 1 = no confidence or belief, and 5 = strong confidence or belief.

experience with depression treatment), and random effects for providers (to address clustering by provider).

To assess the effectiveness of the implementation of the pseudo-cluster randomization, we compared the characteristics of the 2 groups of providers and the observed distribution of number of patients in each arm per provider. To evaluate whether referral bias occurred, we compared the characteristics of the participants randomized to each arm.

To assess the extent of contamination in the usual care arm, we compared key indicators of the depression treatment approach (i.e., use of intervention components) between usual care arm participants being treated by majority-usual care arm providers, usual care arm participants being treated by majority-intervention arm providers, and intervention arm participants (irrespective of provider assignment). Comparison of these 3 groups allows us to distinguish between 3 competing hypotheses. If little contamination occurred with either type of provider, we would expect usual care participants of both majority-usual care and majority-intervention providers to have depression treatment indicators similar to each other, but substantially worse indicators than intervention participants (assuming that high fidelity was achieved in the intervention arm). If contamination of usual care occurred with the majority-intervention providers but was limited with the majority-usual care providers (the goal of pseudo-cluster randomization), we would expect usual care participants of majority-usual care providers to have substantially worse indicators than intervention participants, with indicators for usual care participants of majority-intervention providers falling in the middle. If substantial contamination of usual care occurred with both types of providers, we would expect similar depression treatment indicators in all 3 groups.

Accordingly, we present differences and 95% confidence intervals comparing these indicators between 1) usual care participants of majority-usual care providers versus all intervention participants (expected to favor the intervention) and 2) usual care participants of majority-intervention providers versus usual care participants of majority-usual care providers (in the absence of contamination, expected to be null; in the presence of contamination, expected to favor usual care participants of majority-intervention providers). Differences and confidence intervals were estimated by using generalized linear regression models with an identity link, binomial error distribution, and clustering by provider. Fixed effects for design variables (site, depression treatment experience) yielded convergence problems for the linear-binomial model, given the small cells in some of these subgroup comparisons, and were therefore excluded from these models.

The depression treatment indicators that we considered included the presence and dosing level of an antidepressant at

Downloaded from https://academic.oup.com/aje/article/182/12/1039/2195409 by guest on 25 April 2024

Table 2. Actual Distribution of Participant Arm Assignments, by Provider Assignment, Among HIV Providers and HIV Patients With Depression, SLAM DUNC Study, Southern United States, 2010–2014

Randomization Assignment	No. of Providers	Enrolled Participants			Intervention Participants			Usual Care Participants		
		Total No.	Mean No. per Provider	Range	Total No.	Mean No. per Provider	Range	Total No.	Mean No. per Provider	Range
Majority intervention providers	24[a]	146	6.1	1–18	112	4.7	1–15	34	1.4	0–3
Majority usual care providers	25	158	6.3	1–21	37	1.5	0–5	121	4.8	0–17

Abbreviations: HIV, human immunodeficiency virus; SLAM DUNC, Strategies to Link Antidepressant and Antiretroviral Management at Duke, University of Alabama-Birmingham, Northern Outreach Clinic, and the University of North Carolina at Chapel Hill.

[a] Four providers randomized to majority intervention never enrolled any patients.

the time of enrollment (expected to be balanced) and the presence and dosing level of an antidepressant at 3 months after enrollment (expected to favor the intervention arm). Antidepressant dosing levels were defined as low, moderate, or high as defined previously on the basis of standard dosing guidelines (11, 20). Additional indicators included the proportion

Table 3. Comparison of Baseline Characteristics of HIV Patients With Depression, SLAM DUNC Study, Southern United States, 2010–2014

Characteristic	Intervention ($n = 149$)			Usual Care ($n = 155$)		
	No.	%	Mean (SD)	No.	%	Mean (SD)
Age, years			42.8 (10.3)			44.9 (9.9)
Present sex						
Male	112	75.2		100	64.5	
Female	35	23.5		52	33.5	
Transgender and other	2	1.3		3	1.9	
Sexual orientation						
Heterosexual	58	39.7		81	53.3	
Gay/lesbian	67	45.9		51	33.6	
Bisexual	17	11.6		16	10.5	
Other	4	2.7		4	2.6	
Race/ethnicity						
White non-Hispanic	54	36.2		39	25.2	
Black non-Hispanic	83	55.7		105	67.7	
Hispanic	9	6.0		4	2.6	
Other	3	2.0		7	4.5	
Employment status						
Employed full time	22	15.0		22	14.4	
Employed part time	17	11.6		20	13.1	
Unemployed	108	73.5		111	72.5	
SF-12 physical functioning score (range, 0–100)			44.1 (11.8)			43.8 (12.1)
SF-12 mental functioning score (range, 0–100)			30.5 (9.4)			30.3 (10.4)
No. of HIV symptoms (range, 0–12)			5.2 (2.9)			5.1 (3.1)
Depressive severity (range, 0–50)			20.3 (6.9)			19.9 (6.9)
Self-reported antiretroviral adherence (range, 0–100)			85.8 (23.3)			87.2 (22.2)
CD4 count, cells/mm^3			607.2 (370.9)			569.0 (354.4)
HIV RNA viral load <50 copies/mL	91	68.9		98	67.6	

Abbreviations: HIV, human immunodeficiency virus; SF-12, the Medical Outcomes Study 12-Item Short Form Health Survey; SLAM DUNC, Strategies to Link Antidepressant and Antiretroviral Management at Duke, University of Alabama-Birmingham, Northern Outreach Clinic, and the University of North Carolina at Chapel Hill.

Downloaded from https://academic.oup.com/aje/article/182/12/1039/2195409 by guest on 25 April 2024

of participants on an antidepressant at entry whose dose or medication was changed within 30 days, the proportion of participants not on an antidepressant at entry who received a prescription within 30 days, and the proportion of all participants who received at least 1 antidepressant dose adjustment within 3 months, all expected to favor the intervention arm.

RESULTS

Overall, 53 attending physicians or mid-level providers who managed patients independently provided informed consent and were randomized, with 28 randomized to a majority-intervention arm case mix and 25 randomized to majority usual care (Table 1). The 2 provider groups were balanced with respect to site, training, clinical activity, and confidence and skill in managing antidepressant treatment. Overall, providers dedicated about two-thirds of their clinical effort to HIV care and had been caring for HIV patients for a mean of 14 years.

Of the 49 randomized providers with at least 1 enrolled SLAM DUNC participant, the mean number of participants per provider, including any patients managed in conjunction with fellows, was 6.1 (range, 1–18) and 6.3 (range, 1–21) in the majority-intervention arm and majority-usual care arm groups, respectively (Table 2). Providers in the majority-intervention arm group enrolled an average of 4.7 intervention arm and 1.4 usual care arm participants, while providers in the majority-usual care arm group enrolled an average of 1.5 intervention arm and 4.8 usual care arm participants.

A total of 149 and 155 patients were randomized to the intervention and usual care arms, respectively (Table 3). Participants were mostly between 30 and 55 years of age. The majority were male, black non-Hispanic, and unemployed. Participants had a mean physical functioning score half a standard deviation below the US mean of 50 and a mean mental functioning score 2 standard deviations below the US mean. No strong evidence of referral bias emerged. There were small differences between arms in certain demographic characteristics (intervention arm participants were slightly younger on average and more likely to be Caucasian), but the arms were well balanced on baseline physical and mental health indicators, where referral bias would be expected to be most evident.

Approximately 45% of participants were taking antidepressants prior to study entry, with only a minority taking moderate or high doses (Table 4). This baseline distribution was comparable across all groups. In general, for all follow-up depression treatment indicators, there were pronounced differences between intervention arm and usual care arm participants, but among usual care arm participants, there were no substantive differences between those treated by majority-usual care versus majority-intervention arm providers, suggesting limited contamination of usual care. For example, 81% of intervention arm participants were taking antidepressants at 3 months compared with 56% of usual care arm participants of majority-usual care providers and 59% of usual care arm participants of majority-intervention providers. Among intervention arm participants, 41% received at least 1 dose adjustment within 3 months compared with 15% of usual care arm participants of majority-usual care providers and 12% of usual care arm participants of majority-intervention providers.

In relation to the success of masking, late in the study some providers did comment informally to study investigators that they seemed to have gotten all intervention or all usual care patients, but they did not express any suspicions of imbalance,

Table 4. Assessment of Contamination[a] Among HIV Patients With Depression, SLAM DUNC Study, Southern United States, 2010–2014

Depression Treatment Indicator	Usual Care Arm Participants Treated By:		B vs. A Difference (C)	95% CI	Intervention Arm Participants[b] (n = 149), % (D)	D vs. A Difference (E)	95% CI
	Majority-Usual Care Arm Providers (n = 121), % (A)	Majority-Intervention Arm Providers (n = 34), % (B)					
Antidepressant at baseline							
Any	41.2	41.2	0.0	−20.9, 20.9	47.0	5.8	−7.3, 18.9
Moderate or high dose	15.1	17.6	2.5	−14.3, 19.4	16.8	1.7	−6.9, 10.2
Antidepressant at 3 months							
Any	55.5	58.8	3.4	−18.4, 25.1	81.2	25.7	14.1, 37.4
Moderate or high dose	22.7	20.6	−2.1	−23.0, 18.8	32.9	10.2	1.9, 18.5
If taking antidepressant at baseline: Any adjustment in 30 days?	18.0	7.1	−10.9	−27.3, 5.6	60.0	42.0	27.0, 57.0
If not taking antidepressant at baseline: Any antidepressant prescribed in 30 days?	28.8	26.3	−2.5	−27.7, 22.8	67.1	38.3	20.4, 56.2
All: Any dose adjustment in 3 months?	14.7	12.1	−2.5	−19.1, 14.0	40.9	26.3	16.1, 36.4

Abbreviations: CI, confidence interval; HIV, human immunodeficiency virus; SLAM DUNC, Strategies to Link Antidepressant and Antiretroviral Management at Duke, University of Alabama-Birmingham, Northern Outreach Clinic, and the University of North Carolina at Chapel Hill.

[a] Comparison of depression treatment indicators between usual care arm participants treated by majority-usual care providers (A), usual care arm participants treated by majority-intervention arm providers (B), and intervention arm participants (D).

[b] Regardless of assignment of provider.

Downloaded from https://academic.oup.com/aje/article/182/12/1039/2195409 by guest on 25 April 2024

rather attributing it to the "luck of the draw." At debriefing after the conclusion of the study, 3 study staff who had been involved in recruitment commented that they had observed that certain providers seemed to be getting a lot of intervention arm participants and certain others seemed to be getting a lot of usual care participants.

DISCUSSION

The SLAM DUNC study used pseudo-cluster randomization to balance competing concerns about contamination and referral bias in a randomized trial of a provider-focused depression management intervention for patients in HIV care. The study design and masking appear to have been effective in avoiding referral bias: Had referral bias occurred, differences would have been expected in the baseline characteristics of participants in the 2 arms, but in fact patients in the 2 arms were comparable on all measured baseline physical and mental health indicators. There is also no evidence that majority-usual care arm providers referred with less vigor than majority-intervention arm providers.

The pseudo-cluster design was selected over patient randomization in order to minimize contamination. The present design allows a direct assessment of contamination by comparing depression treatment indicators between usual care arm participants managed by majority-usual care arm providers and usual care arm participants managed by majority-intervention arm providers, relative to the same indicators in the intervention arm. Little evidence of contamination emerged. Usual care arm participants managed by the 2 different types of providers had nearly identical depression treatment indicators, while intervention arm participants had markedly better indicators. Had substantial contamination occurred, we would have expected depression treatment indicators for usual care arm participants of majority-intervention providers to be better (and substantially closer to indicators for intervention arm participants) than indicators for usual care arm participants of majority-usual care providers. It remains possible that some contamination of usual care occurred with both types of providers or that the study was underpowered to identify differences between 2 groups of usual care participants, but the pronounced differences of both groups of usual care participants compared with the intervention arm give strongest support to the hypothesis that no or limited contamination occurred.

The present study masked providers and recruitment staff to the imbalanced randomization design to provide maximal protection against referral bias. Although some providers commented to study investigators that they noticed they had gotten mostly one or the other type of patient assignment, all seemed to attribute the imbalance to the luck of the draw rather than to intentional design. However, masking the imbalance to study staff may have been more challenging, given that the same study staff requested randomizations for all providers at a given site and were therefore more prone to see patterns in the treatment arm assignments. However, to the extent that study staff may have suspected imbalance, it does not appear to have led to referral bias.

At the design stage for this study, the scientific review panel had strong concerns about the threat of contamination in a patient-randomized design, and it seemed likely that such a design would not be funded. At the same time, the investigator team believed that a provider-randomized design would lead to problematic referral bias. Pseudo-cluster randomization provided an avenue to address both concerns simultaneously. Although the weak evidence of contamination suggests after the fact that the pseudo-cluster design may not have been necessary, there was no firm evidence on which to base such an argument at the design stage. On the other hand, the fact that at study exit most providers expressed strong preference for the intervention over usual care suggests that concern about the potential for referral bias in a provider-randomized design was well founded. The use of the pseudo-cluster design provided a unique opportunity to directly evaluate the extent of contamination after the fact, providing invaluable information for future related work. To our knowledge, no other study in this area has assessed actual contamination in this way.

The SLAM DUNC Study is only the second application of pseudo-cluster randomization known to the authors, although suggestions for its usefulness in other settings have been made (21–23). The experiences presented here are consistent with reports from the Dutch EASYCare Study, which found that there was little evidence of suspicion by providers of the imbalanced randomization probabilities (24). The Dutch EASYCare investigators also found that providers had a strong preference for the intervention arm, and that more than half of the providers said at the end of the study that they would likely have recruited fewer patients if all their patients had been assigned to usual care, as would have been the case in a cluster-randomized trial. This finding is in line with other reports of slower or biased recruitment among control clusters in cluster-randomized trials (3, 5–7). Pseudo-cluster randomization could help to address referral bias even if providers and staff are not blinded to the imbalanced randomization, because each participant has a nonzero probability of being assigned to either arm.

With increasing emphasis on and funding for implementation science (25), innovative designs are required that permit the strongest causal inference and protection against bias within the constraints of real-world settings. For the present study, pseudo-cluster randomization appears to have been effective in preventing referral bias while avoiding contamination for a real-world study in which both biases were plausible a priori and would have substantially undermined the randomized design's validity. The design may have applicability in a range of health service delivery studies that face competing threats of contamination and referral bias.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, Gillings School of Global Public Health, the University of North Carolina at Chapel Hill, Chapel Hill, North Carolina (Brian W. Pence); Department of Psychiatry, School of Medicine, the University of North Carolina at Chapel Hill, Chapel Hill, North Carolina (Bradley N. Gaynes); Division of Infectious Diseases, Department of Medicine, School of Medicine, the University of North Carolina at Chapel Hill,

Downloaded from https://academic.oup.com/aje/article/182/12/1039/2195409 by guest on 25 April 2024

Chapel Hill, North Carolina (Amy Heine, Evelyn B. Quinlivan); Division of Infectious Diseases, Department of Medicine, School of Medicine, Duke University, Durham, North Carolina (Nathan M. Thielman); Department of Biostatistics and Bioinformatics, School of Medicine, and Duke Global Health Institute, Duke University, Durham, North Carolina (Elizabeth L. Turner); and Division of Infectious Diseases, School of Medicine, the University of Alabama at Birmingham, Birmingham, Alabama (Michael J. Mugavero).

This study was supported by grant R01MH086362 of the National Institute of Mental Health and the National Institute for Nursing Research, Bethesda, MD. Support was also provided by the Centers for AIDS Research at the University of North Carolina at Chapel Hill, Duke University, and the University of Alabama at Birmingham National Institutes of Health-funded programs (P30-AI50410, P30-AI064518, and P30-AI027767).

We are indebted to Anne Fletcher who identified pseudo-cluster randomization during the design phase of the SLAM DUNC Study.

The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health, the National Institute for Nursing Research, or the National Institutes of Health.

Conflict of interest: none declared.

REFERENCES

1. Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. London, UK: Arnold (Hodder Headline Group); 2000.
2. Torgerson DJ. Contamination in trials: Is cluster randomisation the answer? *BMJ*. 2001;322(7282):355–357.
3. Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *BMJ*. 2003;327(7418): 785–789.
4. Puffer S, Torgerson DJ, Watson J. Cluster randomized controlled trials. *J Eval Clin Pract*. 2005;11(5):479–483.
5. Jordhøy MS, Fayers PM, Ahlner-Elmqvist M, et al. Lack of concealment may lead to selection bias in cluster randomized trials of palliative care. *Palliat Med*. 2002;16(1):43–49.
6. Eldridge S, Kerry S, Torgerson DJ. Bias in identifying and recruiting participants in cluster randomised trials: What can be done? *BMJ*. 2009;339:b4006.
7. Hahn S, Puffer S, Torgerson DJ, et al. Methodological bias in cluster randomised trials. *BMC Med Res Methodol*. 2005;5:10.
8. Donner A, Birkett N, Buck C. Randomization by cluster. Sample size requirements and analysis. *Am J Epidemiol*. 1981; 114(6):906–914.
9. Teerenstra S, Melis RJ, Peer PG, et al. Pseudo cluster randomization dealt with selection bias and contamination in clinical trials. *J Clin Epidemiol*. 2006;59(4):381–386.
10. Pence BW, Gaynes BN, Williams Q, et al. Assessing the effect of Measurement-Based Care depression treatment on HIV medication adherence and health outcomes: rationale and design of the SLAM DUNC Study. *Contemp Clin Trials*. 2012; 33(4):828–838.
11. Adams JL, Gaynes BN, McGuinness T, et al. Treating depression within the HIV "medical home": a guided algorithm for antidepressant management by HIV clinicians. *AIDS Patient Care STDS*. 2012;26(11):647–654.
12. Teerenstra S, Moerbeek M, Melis RJ, et al. A comparison of methods to analyse continuous data from pseudo cluster randomized trials. *Stat Med*. 2007;26(22):4100–4115.
13. Hayes RJ, Moulton LH. *Cluster Randomised Trials*. Boca Raton, FL: Chapman & Hall/CRC Press; 2009.
14. Safren SA, O'Cleirigh C, Tan JY, et al. A randomized controlled trial of cognitive behavioral therapy for adherence and depression (CBT-AD) in HIV-infected individuals. *Health Psychol*. 2009;28(1):1–10.
15. Pals SL, Beaty BL, Posner SF, et al. Estimates of intraclass correlation for variables related to behavioral HIV/STD prevention in a predominantly African American and Hispanic sample of young women. *Health Educ Behav*. 2009;36(1): 182–194.
16. Parker DR, Evangelou E, Eaton CB. Intraclass correlation coefficients for cluster randomized trials in primary care: the Cholesterol Education and Research Trial (CEART). *Contemp Clin Trials*. 2005;26(2):260–267.
17. Smeeth L, Ng ES. Intraclass correlation coefficients for cluster randomized trials in primary care: data from the MRC Trial of the Assessment and Management of Older People in the Community. *Control Clin Trials*. 2002;23(4):409–421.
18. 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.
19. Bess KD, Adams J, Watt MH, et al. Providers' attitudes towards treating depression and self-reported depression treatment practices in HIV outpatient care. *AIDS Patient Care STDS*. 2013;27(3):171–180.
20. Landis SE, Gaynes BN, Morrissey JP, et al. Generalist care managers for the treatment of depressed Medicaid patients in North Carolina: a pilot study. *BMC Fam Pract*. 2007;8:7.
21. Vaucher P. Designing phase III or IV trials for vaccines: choosing between individual or cluster randomised trial designs. *Vaccine*. 2009;27(13):1928–1931.
22. de Jong K, Moerbeek M, van der Leeden R. A priori power analysis in longitudinal three-level multilevel models: an example with therapist effects. *Psychother Res*. 2010;20(3): 273–284.
23. Klonoff DC, Bergenstal R, Blonde L, et al. Consensus report of the Coalition for Clinical Research—Self-Monitoring of Blood Glucose. *J Diabetes Sci Technol*. 2008;2(6):1030–1053.
24. Melis RJ, Teerenstra S, Rikkert MG, et al. Pseudo cluster randomization performed well when used in practice. *J Clin Epidemiol*. 2008;61(11):1169–1175.
25. Meissner HI, Glasgow RE, Vinson CA, et al. The U.S. training institute for dissemination and implementation research in health. *Implement Sci*. 2013;8:12.

Downloaded from https://academic.oup.com/aje/article/182/12/1039/2195409 by guest on 25 April 2024