

Randomization, design and analysis for interdependency in aging research: no person or mouse is an island

Received: 20 June 2022

Accepted: 31 October 2022

Published online: 22 December 2022

 Check for updates

Daniella E. Chusyd¹, Steven N. Austad^{2,3}, Stephanie L. Dickinson⁴, Keisuke Ejima⁴, Gary L. Gadbury⁵, Lilian Golzarri-Arroyo⁴, Richard J. Holden⁶, Yasaman Jamshidi-Naeini⁴, Doug Landsittel⁴, Tapan Mehta⁷, J. Michael Oakes⁸, Arthur H. Owora⁴, Greg Pavela⁹, Javier Rojo⁴, Michael W. Sandel¹⁰, Daniel L. Smith Jr.^{3,11}, Colby J. Vorland¹², Pengcheng Xun^{4,13}, Roger Zoh⁴ & David B. Allison⁴✉

Investigators traditionally use randomized designs and corresponding analysis procedures to make causal inferences about the effects of interventions, assuming independence between an individual's outcome and treatment assignment and the outcomes of other individuals in the study. Often, such independence may not hold. We provide examples of interdependency in model organism studies and human trials and group effects in aging research and then discuss methodologic issues and solutions. We group methodologic issues as they pertain to (1) single-stage individually randomized trials; (2) cluster-randomized controlled trials; (3) pseudo-cluster-randomized trials; (4) individually randomized group treatment; and (5) two-stage randomized designs. Although we present possible strategies for design and analysis to improve the rigor, accuracy and reproducibility of the science, we also acknowledge real-world constraints. Consequences of nonadherence, differential attrition or missing data, unintended exposure to multiple treatments and other practical realities can be reduced with careful planning, proper study designs and best practices.

Investigators traditionally use randomized trials, or experiments, and corresponding analysis to make causal inferences about the effects of interventions, assuming independence between an individual's outcome and treatment assignment and other individuals' outcomes in the

study. In aging research, however, this assumption of independence is not always valid. Examples of interdependency include interference¹, group composition effects² and clusters and nesting³. These issues require attention because they may violate the assumptions of causal

¹Department of Environmental and Occupational Health, Indiana University-Bloomington, Bloomington, IN, USA. ²Department of Biology, University of Alabama at Birmingham, Birmingham, AL, USA. ³Nathan Shock Center, University of Alabama at Birmingham, Birmingham, AL, USA. ⁴Department of Epidemiology and Biostatistics, Indiana University-Bloomington, Bloomington, IN, USA. ⁵Departments of Statistics, Kansas State University, Manhattan, KS, USA. ⁶Department of Health & Wellness Design, Indiana University-Bloomington, Bloomington, IN, USA. ⁷Department of Family and Community Medicine, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. ⁸Department of Quantitative Health Science, Case Western Reserve University, Cleveland, OH, USA. ⁹School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA. ¹⁰Department of Wildlife, Fisheries and Aquaculture, Mississippi State University, Starkville, MS, USA. ¹¹Department of Nutrition Sciences, University of Alabama at Birmingham, Birmingham, AL, USA. ¹²Department of Applied Health Science, Indiana University-Bloomington, Bloomington, IN, USA. ¹³Department of Global Value, Access and Outcomes, Atara Biotherapeutics, Thousand Oaks, CA, USA. ✉e-mail: allison@iu.edu

inference and of independence made when using traditional hypothesis tests. These terms and others are often not defined uniformly, however, which can lead to confusion. For the purpose of this report, we have defined a set of terms in Box 1.

Interdependency has begun to be addressed in the scientific literature^{4–7} but has received little attention in aging research. Yet, the interdependence of subjects within subject-clusters can be observed in the designs and analyses of aging studies. Although the field acknowledges that it is difficult to disentangle how the nine recognized hallmarks of aging are connected⁸, these undoubtedly impact one another and may themselves be sources of or characterized by interdependency.

These study design challenges underscore the importance of the National Institute on Aging's effort to 'develop innovative changes in the design, planning and implementation of clinical trials'⁹. Indeed, aging research requires researchers to address interdependency through proper study design, analysis and interpretation (Table 1 and Fig. 1). In this Perspective, we highlight the use and importance of randomization and summarize examples of interdependency and related methodologic issues to call attention to interference, clustering and independence and significance levels in aging research (Box 2).

Examples of interdependency in aging research

Statistical interdependence in animal models

A ubiquitous issue in experimental paradigms using the three main animal models in aging research—*Caenorhabditis elegans* (hereafter 'worms'), *Drosophila melanogaster* (hereafter 'flies') and mice—is housing animals in multiple separate enclosures but combining results as if the animals formed a single population. In worms, survival studies generally combine data from subpopulations maintained on multiple agar plates or multiple wells for liquid culture. For instance, the *C. elegans* Interventions Testing Program, which has extensively explored the replicability of lifespan studies among laboratories¹⁰, uses at least three agar plates each containing 35 to 40 animals to complete a single survival assay. Other studies use as few as 20 to 30 individuals per plate and combine the results of several plates^{11,12}. Surprisingly, the number of plates or vials involved in survival analysis is often not specified. In any case, individual plates have a separate history and micro-environment, varying density over time as animals die, and possibly different personnel transferring animals to fresh plates. The important impact of precise transfer technique on longevity has been established by the *C. elegans* Interventions Testing Program.

Similarly, fly researchers use a wide variety of housing conditions (for example, cages, bottles and vials) but most typically combine survival results from 5 to 10 vials each containing 20 to 30 flies¹³ nearly always separated by sex, because mixed-sex housing is known to shorten the lives of both sexes^{14,15}. Some studies use substantially larger samples and cages, for instance 125 flies in 3 to 5 replicates, but typically combine replicates for the demographic analyses¹⁶. As with worms, each fly vial will have its individual history and micro-environment and possibly different personnel transferring flies to fresh enclosures periodically.

Mouse studies, in which the phenotype of individuals is more easily studied than in worms or flies, typically house four mice or fewer per cage with sexes separated at the beginning of survival experiments, although some research suggests that short-term health is not compromised by higher densities¹⁷. Male mice are often from the same litter to minimize fighting, but fighting among males is a recurring issue, resulting in individual males, or even whole cages, being removed from studies¹⁸. The number of animals housed in a cage alters thermal and social environments, affecting organ weight, heart rate and multiple aspects of behavior, including food consumption and torpor (particularly important because torpor may be associated with the longevity benefit of food restriction)^{19,20}. Nearly all animal facilities are maintained at temperatures markedly below rodent thermoneutrality²¹. Group-housed animals somewhat compensate

for this by huddling. The impact of the thermal environment can easily be seen when mice or rats are housed singly. In one study, singly housed mice ate 40% more than mice housed in groups of four while maintaining similar body weights²². The thermal environment also affects body composition, the ratio of brown-to-white fat²³, activity, and, over time, pathology^{24,25}. Group-housed mice also display greater phenotypic variability than singly housed mice²⁶. For aging studies, these issues are particularly germane because density will change over time as animals begin to die.

Human trials and group effects

Groups exert substantial influence on the behaviors and outcomes of individuals. A classic example of group effects is the Asch conformity experiments, which demonstrated individuals have a tendency to 'conform' to an erroneous group consensus²⁷, and have been studied for differential patterns with aging. Specifically, older people demonstrate lower rates of social conformity compared with younger individuals²⁸. Another example is of socially induced stress, which can negatively affect longevity in various social species, including humans^{2,29,30}. Despite the intuitive influence of group effects, the rigorous identification of group effects per se, also called peer effects or contagion effects, is difficult^{31,32}. This is particularly relevant in aging-related research involving older persons in congregate settings. Such circumstances by their nature tend to involve interdependency and examples of studies involving cluster-randomized trials^{33,34}, pseudo-cluster randomization^{35,36}, group composition designs³⁷ and individually randomized but group-delivered trials^{38,39} exist. For example, herd immunity can affect the analysis of vaccine efficacy⁴⁰, as discussed in 'Cluster-randomized controlled trials'. The ACTonHEART intervention⁴¹ is another example of potential group effects. In that study, individuals (not clusters) were randomly assigned but received the intervention in group-therapy sessions (that is, post-randomization clustering)⁴¹. Another more subtle example of a potential group effect occurs when individuals share an interventionist. For example, the Dutch Geriatric Intermediate Care Program was designed to assess the effect of home visits by geriatric nurses on the function of older adults compared with usual care. Older adults shared their general practitioners. Thus, the general practitioner's exposure to those in the intervention group could affect the care provided to the usual care group. In trials in which a treatment is administered in a group setting or a single interventionist administers a study intervention to multiple participants, we can observe both interference and within-group correlation of outcomes because of group composition effects.

In observational studies of contagion effects, the challenges are compounded because of homophily and shared environment⁴². Confounding due to homophily occurs when the same factor that influences an individual's outcome of interest also influences that individual's propensity to form ties (and the strength and duration of ties) with others characterized by the exposure of interest. Environmental confounding occurs when an individual and a group share an environmental factor associated with the outcome of interest. In either homophily or environmental confounding, it is difficult to disentangle the causal effect of one's peers from shared peer characteristics and environmental characteristics shared with one's peers. One area in which this may occur is when studying centenarians, who are often studied for insight into long, healthy lives. If a study design focuses on identifying 'longevity genes' within certain families, for example, issues of interdependence associated with shared environments are raised^{43,44}. Other issues associated with exceptional longevity are age-cohort effects, for instance, among those born before, during or after major environmental or political events (for example, war or pandemic)^{45,46}.

Group formation experiments, in which individuals are randomly assigned to groups of varying compositions and an outcome of interest is observed, can overcome some of the limitations inherent to

BOX 1**Key concepts**

- **Cluster.** A socially intact unit (for example, nursing home, family, hospital and community) in which individuals are naturally grouped in most cases. 'In cRCTs, observations within a cluster are likely to be more similar than observations in other clusters.'⁸⁰
- **cRCT.** Also called a grouped randomized trial, denotes an experiment in which grouped individuals in socially intact units (for example, community, workplace, nursing home and family) are randomly assigned to different levels of the independent variable (for example, a lifestyle intervention program for smoking cessation)^{93,94}.
- **Contamination.** 'The phenomenon of contamination is also variously referred to as leakage⁹⁵, spillover effects⁹⁶ or treatment diffusion⁹⁷. Contamination occurs when interaction between individuals randomly assigned to different treatment conditions causes some individuals to receive features of a treatment to which they were not assigned'⁹⁸.
- **Direct causal effect.** The direct effect of a treatment on an individual as the difference between the potential outcome for that individual given treatment compared with the potential outcome for that individual without treatment, all other things being equal. On the population-average level, it compares potential outcomes of individuals allocated to treatment in treatment clusters with the potential outcomes of individuals allocated to control in treatment clusters.
- **Effectiveness.** The intervention effect under usual condition of care.
- **Efficacy.** The intervention effect under ideal conditions.
- **Environmental confounding.** Occurs when an individual and the peers or group presumed to possess a characteristic that exerts a causal influence on that individual share an environmental factor associated with the outcome of interest.
- **Experiment.** A study in which experimental units (for example, mice and people) are randomly assigned to different levels of the independent variable (for example, a therapeutic intervention).
- **Experimental unit.** A unit in the study that can independently be assigned treatment, thus creating a 'true replicate' (true replicates are not always easily discerned or obtained; for example, a cage of housed mice at a particular temperature setting would represent a true replicate, and the individual mice are considered correlated and have been referred to as pseudoreplicates³³) in a randomized experiment.
- **External effect.** Synonymous with interference. Occurs when the outcome of a given individual is affected by the treatment assignments of other individuals⁸⁹.
- **Group composition effects.** Additional effects, over and above the effects of an individual's characteristics, found when those individual characteristics are aggregated at a higher level, such as a group or cluster.
- **Homophily.** The tendency of individuals to associate with similar others.
- **Intention-to-treat analysis.** Analyzing participants according to how they were originally randomized, even if they did not complete the study.
- **Interference.** Denotes a phenomenon that the exposure or treatment received by one individual may affect the outcomes of other individuals¹.
- **ICC coefficient.** 'The most common way of quantifying the extent of nonindependence as a function of clustering'⁸⁰.
- **Nesting.** 'In cRCTs, the nested or hierarchical structure of the design changes the degrees of freedom for testing the intervention effect because 'the units of observation are nested within the units of (randomization)' (ref. 99). The degrees of freedom depend both on the number of units of observation and on the number of units of randomization'⁸⁰.
- **Overall causal effect.** The average effect of an intervention relative to no intervention. It compares the potential outcomes of all individuals in clusters allocated to treatment with those of all individuals in clusters allocated to control.
- **Potential outcomes.** The outcomes of an experimental unit that are potentially observable in a study; each experimental unit has a potential outcome corresponding to each possible treatment to be assigned. In practice, only one outcome is observable at a particular time depending on the treatment that the experimental unit actually received. The other unobservable outcomes are referred to as counterfactual¹⁰⁰.
- **Pragmatic design.** Implementation of an intervention as it would work in practice.
- **Pseudo-cluster randomization.** A special case of two-stage randomization. In the first stage, the clusters are randomized into two groups. In the second stage, in one group of clusters, most of the individuals or participants (for example, 80%) will randomly receive the treatment, and in the other group of clusters, the majority will randomly receive the control condition⁸¹.
- **Randomized trial.** Synonymous with experiment and usually reserved for an experiment with human participants randomly assigned to different levels of the independent variable (for example, a medical or educational treatment).
- **Spillover effect.** A form of interference and describes the effect on an individual of the treatment received by others in the group. On the population-average level, it compares potential outcomes of individuals allocated to control in treatment clusters with the potential outcomes of individuals allocated to control in control clusters.
- **Stable unit treatment value assumption.** The assumption that 'there is no interference between units (ref. 101) leading to different outcomes depending on the treatments other units received and there are no versions of treatments leading to technical errors (ref. 102,103)'.⁶
- **Stratified interference.** The assumption that there is no interference across clusters. It is also referred to as the 'partial interference assumption' because it could be viewed as an intermediate assumption between (1) assuming no interference within a group and (2) making no assumptions about the nature of interference within a group⁶.
- **Total causal effect.** Describes both the direct and indirect effects of a particular treatment assignment on an individual. On the population-average level, it compares potential outcomes of individuals allocated to treatment in treatment clusters with the potential outcomes of individuals allocated to control in control clusters.
- **Two-stage randomization.** Randomly allocating treatments across level-2 units (for example, communities) and randomizing the treatments themselves across individuals (level-1 units) within level-2 units^{6,89,91}.

Table 1 | An overview of the discussed six study designs

Design/topic/circumstance	Randomization applied to:	Treatment applied to:	Measurement taken on:	Advantages/opportunities	Challenges	Comments
(1) Group composition experiments	Individual level	Clusters or groups of individuals	Individual level	Estimate the effects of a group characteristic (for example, average age) on individual outcomes.	Model identification of peer effects.	Randomization helps to address, but may not eliminate, problems of reflection.
(2) Single-stage individually randomized trials	Individual level	Individual unit	Individual level	When assumptions hold, straightforward to design, analyze and interpret results. When individual units are randomized and remain independent and connection among groups is reduced or eliminated, interference is also minimized.	In human trials, adherence, dropout, patient and physician preferences, and it is impossible to assess interference.	Some analysis methods such as intent-to-treat analyses can address some challenges but may still have limitations with time-varying covariates or moderators.
(3) Cluster-randomized controlled trials	Clusters or groups of individuals	Cluster or groups of individuals	Cluster or groups of individuals	Can obtain more precise treatment comparison compared to the single-stage individual randomization under some circumstances. Contamination of one treatment to the other is minimized by the groups being separately assigned.	Outcomes are correlated within the cluster and requires estimation of the ICC for valid analysis. Statistical power may be low if the number of clusters is small. Effect of treatment may include direct effects as well as indirect effects (for example, herd immunity in vaccine trials).	Clustering and nesting are important considerations in the analyses of data, and degrees of freedom are based on the number of clusters, not the number of individuals within the cluster.
(4) Pseudo-cluster-randomized trials	Clusters or groups first to ‘intervention majority’ (H) or ‘control majority’ (L), then a fraction f ($0.5 \leq f \leq 1$) of the individuals within H and L clusters randomized to treatment and control, respectively.	Individual level	Individual level	Rate of contamination is expected to be reduced compared to individually randomized designs.	Contamination may still occur when the treatment is easy to execute and/or when the numbers assigned to each of the two interventions within a group are large. Perfect blinding may be difficult if study personnel are able to predict the treatment assignments over time.	Can assess contamination by comparing the treatment effect among minority control, majority control and intervention subjects (minority and majority inclusive).
(5) Individually randomized group treatment	Individual level	Clusters or groups of individuals	Individual level	Can observe both interference and within-group correlation of outcomes.	Nonindependence in observations that need to be accounted for, similar to the cRCT, and estimation of the ICC may be more complicated.	Correlations may develop as group members share a treatment environment.
(6) Two-stage randomized designs	Nonoverlapping groups first, followed by units in the group	Units in the group	Units in the group	May allow for separate estimation of direct versus indirect effects, and this allows identification of interference.	Harder to implement than a traditional RCT or group trial. Compliance can be an issue.	There are variations of two-stage designs with varying advantages and challenges to proper implementation.

observational studies⁴⁷. The goal of randomized group formation experiments is to isolate the causal effect of a group characteristic on individual outcomes. However, the random assignment of individuals to groups does not resolve the problem of confounding due to shared environments⁴⁸. Nor does random assignment to a peer group guarantee the random formation of network ties. Given the challenges of isolating peer effects on individual outcomes, statistical methods for the estimation of peer effects—both in randomized and nonrandomized designs—is an active area of development and discussion. A common method for estimating peer effects is the linear-in-means model, in which the outcome of interest is regressed on an individual’s characteristics and the average peer outcomes and characteristics^{47,49}. Sacerdote⁵⁰ provides a thorough review of a peer-effects linear-in-means model, including its limitations, and other approaches to estimate and identify peer effects in group composition experiments.

Real-world constraints and recommendations

Trial recruitment in naturalistic settings is subject to the challenges described throughout this Perspective. This is especially true in pragmatic trials with human participants. A well-known difficulty in clinical trials involves whether people comply with their assigned treatment or remain in the study until its completion. If the person does not comply or leaves the trial, the study contains missing data, and much has been written on this issue⁵¹. For instance, trials of technology interventions suffer from systematic and cumulative nonadherence and attrition in the treatment arm⁵², a phenomenon that may be more common in subgroups affected by a ‘digital divide’, such as rural participants⁵³ and older adults⁵⁴. In these trials, nonadherence, differential attrition or missing data, unintended exposure to multiple treatments, and other practical realities occur probabilistically but not inevitably; certain study designs and best practices can reduce the risk and consequence of these effects.

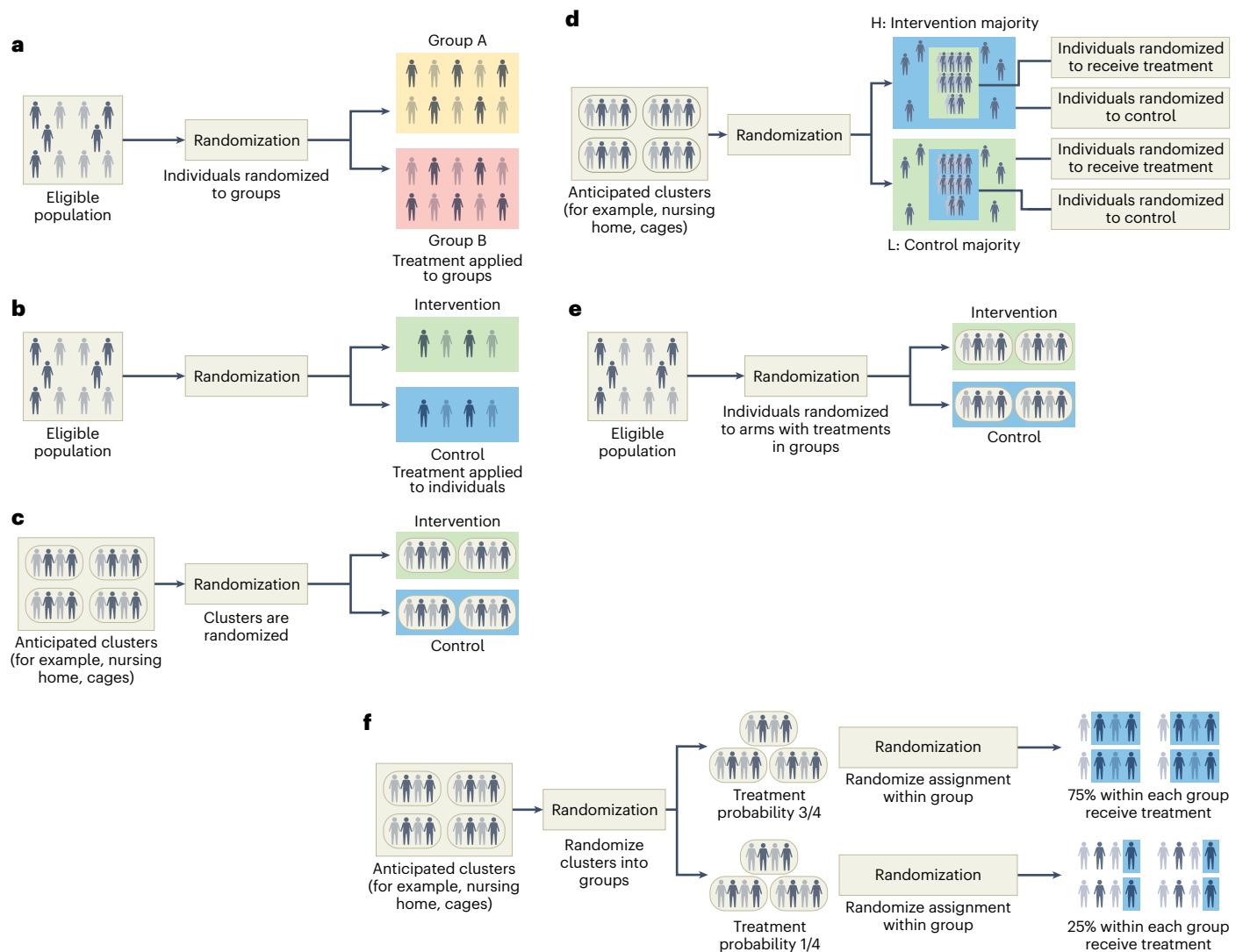


Fig. 1 | A visual representation of the study designs. a, Group composition experiments. **b**, Single-stage randomized trials. **c**, cRCTs. **d**, Pseudo-cluster-randomized trials. **e**, IRGT. **f**, Two-stage randomized designs.

The intention-to-treat effect can still be estimated to evaluate the effect of being randomized to a given condition even if participants do not complete the study³⁵. While sometimes criticized, the intention-to-treat analysis serves a valuable purpose from a public health perspective: the effect of random assignment on the population. In this way, investigators can assess whether use of a guideline, policy or other intervention has a significant effect versus not implementing the (or implementing a different) guideline, policy or intervention. Although effectiveness from the public health perspective does not properly estimate efficacy, or even effectiveness from the patient perspective, it does inform policy, public health and clinical decision-making, which are particularly important in aging research.

A related but different issue is assessing the utility of using a pragmatic design for a given research question. The answer to this question relates to, in large part, whether the intervention dose is sufficiently different in the intervention arm versus control arm. For instance, if the pragmatic study is assessing whether care facilitated by physician alerts affects health, the physician alerts must reach a sufficiently larger percentage of participants in the intervention arm to even assess the intervention effect. Otherwise, results are likely to be nonsignificant even if the intervention itself is effective. Further, overlap between the arms may be greatly affected by the experimental unit and other

interdependencies. By contrast, a pragmatic trial may be necessary when the results of a traditional randomized controlled trial (RCT) are not generalizable. For instance, if persons of lower socioeconomic status are highly underrepresented in the trial, that sampling procedure will greatly affect the utility of the findings.

Although nonadherence, differential attrition or missing data, unintended exposure to multiple treatments, and other practical realities frequently occur, they are not inevitable. Careful planning, proper study designs and best practices can reduce the risk and consequence of these occurrences. Research teams can perform a risk assessment of any potential threats to valid inference at the outset of the study and have clear and detailed protocols in place to mitigate anticipated challenges. When unforeseen issues arise, resultant contamination, nesting and other interdependencies can often be measured and accounted for in analysis. If nothing else, deviations from protocol should be documented clearly to allow for accurate and transparent reporting.

Available study designs

Single-stage individually randomized trials

In a single-stage individually randomized trial, a control group is expected, in probability, to be identical to the intervention group at baseline. That is, the average attributes of the two groups are assumed

BOX 2

Illustrations of interdependency in geroscience and aging research

- A study investigated the incidence of coronavirus disease 2019 (COVID-19) in residents and staff of a nursing facility when participants were randomized to receive bamlanivimab or a placebo. However, COVID-19 incidence in the control group is interdependent with the treatment group. People in the treatment group are now less likely to develop COVID-19, thus there are fewer people who can infect individuals in the control group¹⁰⁴.
- A study investigated whether weight cycling (that is, repeated weight gain followed by weight loss) altered lifespan in mice. However, the treatment is entirely correlated with the cage as all mice in cage 1 received treatment 1, all mice in cage 2 received treatment 2 and all mice in cage 3 received treatment 3 (ref. 67).
- In some social species, population dynamics (for example, lifespan and fitness) are influenced by Allee effects, which are manifestations of the nonlinear relationship between population density and individual fitness¹⁰⁵. Heat (energy) conservation among group-housed mice is a clear example of an Allee effect and can lead to an interdependent outcome.
- Various forms of communication influence animal social structure and behavior. In the eusocial naked mole rat, subordinate colony members consume the feces of the queen (the dominant female) and engage in alloparental behaviors, suggestive of pheromone communication, while also displaying delayed or incomplete reproductive maturity. Coprophagy and downstream physiologic effects are rarely measured, and the degree to which pheromones affect behavior and social structure of experimental colonies remains an important source of potential interference¹⁰⁶.
- A study demonstrated that larval population density impacts the developmental rate and adult lifespan of *C. elegans*. This is an example of how study design may lead to interdependency in the outcome of interest¹⁰⁷.
- A study provided evidence that in older adults, the risk of a major cardiovascular event is increased in the immediate weeks following the loss of a spouse¹⁰⁸. This example illustrates how potentially unrelated or unaccounted-for factors in a study design can impact the outcome.
- A study demonstrated that dietary knowledge improved in older adults who do not live alone¹⁰⁹, providing evidence of potential interdependency in the outcome. Information one person is receiving may spill over to other people in the household.

to be the same. Therefore, statistically significant differences in the outcome can be attributed to the intervention. When baseline covariates are suspected to influence outcomes in a systematic way (for example, participant age in a survival analysis, disease severity, offspring of animal models being measured from successive progeny (for example, F_1 , F_2 , F_3 and F_4) versus from different parity⁵⁶), covariate considerations and adjustments may be useful at the design (for example, stratified randomization⁵⁷) and analysis (for example, randomization-based⁵⁸ and model-based analysis⁵⁹) stages, respectively.

In parallel-group efficacy RCTs, the power to detect statistical interactions between treatment and baseline strata is often low

compared with the power to evaluate an average treatment effect. For example, the lack of evidence for treatment efficacy among women and men based on separate analyses does not address the question of whether treatment differences vary depending on sex⁶⁰. Moreover, multiple subgroup analyses involving baseline strata like age or disease stage or multiplicity involving analysis of several endpoints can increase type 1 error rates. Conversely, correction for such errors (that is, multiple comparisons adjustment or multiplicity adjustment) may increase type 2 error rates. Thus, tests of exploratory or confirmatory interaction hypotheses should precede within-subgroup analysis.

In existing aging-related trials, most intention-to-treat analyses rely exclusively on comparison of baseline treatment assignment to determine treatment effectiveness and ignore potential time-varying covariate issues^{61,62}. But time-varying covariates, in other words, prognostic factors that change, can result in changes in the treatment or intervention over time, which in turn affect treatment efficacy measures. Identifying potential time-varying covariates is important to understand the causal effects of investigated treatments or interventions⁶³.

Potential time-varying moderators must also be considered^{64,65}. These include factors that may change over time (including measuring the outcome⁶⁶) and modify the treatment effect on outcomes of interest, including breeding strategies or ‘cohort’ effects. Additional factors that may change over time include secondary mutations resulting from genetic drift.

Cluster-randomized controlled trials

A cluster-randomized controlled trial (cRCT) is a trial in which the randomization units are clusters or groups of individuals (for example, clinics, hospitals, classes and families) instead of individuals themselves, although outcomes are measured at the individual level. In this case, the outcomes are likely to be correlated within the cluster and are not independent observations as is the assumption of standard statistical analyses such as *t*-tests, analysis of variance or regression as typically used.

There are two important issues with this design: clustering and nesting. Clustering means that individuals are grouped together (for example, patients within a clinic or mice within a litter). Nesting means that clusters or groups are situated within a treatment regimen such that all individuals in the same cluster receive the same treatment. For example, in the study by List et al.⁶⁷, mice were clustered within the cage, and cages were nested within the treatment because all mice in the same cage received the same diet. Clustering is measured by the intraclass correlation (ICC), which describes the amount of the variation of the data explained by the unit of randomization (that is, the cluster)⁶⁸, meaning the correlation within clusters relative to the correlation between clusters. Ignoring clustering and nesting during analyses can lead to an inflated type I error rate^{3,69–72}. There are additional issues, such as census recruitment or enrolling via cluster random sampling, a two-stage process in which the population is divided into clusters and a subset of the clusters is randomly selected, as opposed to investigator-led selection of clusters, which can be argued to induce bias and we refer the reader elsewhere for detailed discussions^{73–75}.

Additionally, because clusters are the independent unit of analyses, the analysis needs to account for the number of clusters, the ICC, and the number of individuals per cluster. When the number of clusters is small, and the coefficient of variation is even moderately large⁷⁶, statistical power to detect treatment effects will be limited regardless of the sample size within clusters^{70,71}. It is important to correctly specify the degrees of freedom according to the independent units of randomization.

Even when clustering is carefully considered, individuals in the same cluster may interfere with each other, such that the estimated (direct) effect may be biased (we use the word ‘bias’ several times; whether a procedure is biased depends in part on the estimand⁷⁷).

For example, when a cRCT is used to estimate a vaccine's effect (where clusters are assigned to vaccine or placebo), vaccine efficacy tends to be overestimated when using a typical approach for analyzing cRCT data. This occurs because the estimated vaccine efficacy reflects the vaccine's direct and indirect effects, and those two effects cannot be distinguished by comparing vaccinated and unvaccinated individuals. Indirect effects appear as the result of herd immunity, where individuals in the vaccinated group are exposed to fewer pathogens because others in the community are also vaccinated⁴⁰. Thus, the magnitude of exposure to a pathogen is correlated within clusters. To identify an effective vaccine, such overestimation may erroneously appear to be beneficial due to the high power. A simulation study demonstrated that disease contagiousness creates a high ICC; thus, any perceived benefit of overestimating the vaccine efficacy in power is diminished⁷⁸. Ultimately, when performing and analyzing a cRCT it is important to collect and analyze the data with a study design and statistical model that accounts for both the ICC (to adjust the denominator degrees of freedom to account for the independent unit of analyses) and the problem of interference. Information on how to analyze this design^{68,69,71,79}; guidelines to follow when describing, analyzing and performing a cRCT⁷⁰; and information to help guide the editorial and peer review process when reviewing cRCTs⁸⁰ can be found in the cited literature.

Pseudo-cluster-randomized trials

As described above, in some studies an individual's initially random treatment assignment may be influenced by the treatment status of other units within a cluster, resulting in a possibly inflated type I error rate. One approach to avoid such contamination (that is, spillover effects) is a cRCT. However, when cRCTs are not possible, or may introduce bias, pseudo-cluster randomization can be considered. Pseudo-cluster randomization is a compromise between cRCT and individual randomization and may be used when there is risk for contamination with randomizing individuals and concern regarding selection bias with randomizing clusters⁸¹.

Pseudo-cluster randomization is a specific type of two-stage randomization⁸² (detailed later in the paper), in which clusters are first randomized to groups labeled H (intervention majority) and L (control majority; more than two groups could be used). In the second step, a fraction f ($0.5 \leq f \leq 1$) of the individuals within H clusters are randomly assigned to treatment and the rest to control. In L clusters, the same fraction f of individuals in each cluster are randomized to control and the rest to treatment⁸². Compared with cluster randomization, selection bias is less likely to arise in pseudo-cluster randomization because the study personnel do not know to which type of cluster (that is, H or L) individuals have been assigned nor do they know (as opposed to cluster randomization) to which treatment a participant will be assigned. However, predictability of treatment assignment would still be an issue with pseudo-cluster-randomized designs. Study personnel might be able to guess the treatment assignments over time with increasing precision, which reintroduces the risk for selection bias. Smaller f fractions will result in lower predictability³⁵.

Reducing contamination in pseudo-cluster randomization (as opposed to individual randomization) is predicated on two underlying assumptions. First, limiting cross-exposure to the other condition reduces contamination. The closer f is to 1, the less the majority condition in each cluster is contaminated by the minority condition. Second, contamination of the majority condition by the minority condition in the same cluster is smaller than vice versa. Whether these assumptions hold depends on the cluster size and the nature of the intervention.

An indirect approach to assessing the extent of contamination in a pseudo-cluster-randomized design is to compare the treatment effect among minority control, majority control, and intervention individuals (minority and majority inclusive). The assumption is that if contamination is small, the treatment effect would be similar in the minority control and the majority control, and substantially smaller

in both control groups compared with the intervention group⁸³. While pseudo-cluster randomization is tagged as a design to reduce contamination, selection bias and recruitment issues of individual and cluster randomizations, there is not a feasible approach to quantify the reduction of contamination by this design compared with individual and cluster randomizations.

Individually randomized group treatment

In individually randomized group treatment (IRGT) trials, individuals are randomly assigned to study conditions. However, unlike in single-stage individually randomized trials, individuals in IRGT trials receive whole or part of their intervention in a group setting. IRGT trials are also in contrast to group randomized trials, which randomly assign clusters and not individuals to study conditions. IRGT trials could involve at least one of the following: (1) individuals in one arm only (typically the intervention) receive treatment in a group setting; (2) individuals in all study arms are administered treatment in a group setting; (3) part of the intervention is administered in a group format; and (4) the intervention is provided by a common interventionist. IRGT trials in which participants in one arm are administered a group intervention are also referred to as partially clustered or partially nested designs^{84,85}. These situations often occur in studies with behavioral components such as exercise or weight loss interventions, which may be delivered in group settings³⁸. For example, the 'Calorie Restriction in Overweight SeniorS: Response of Older Adults to a Dieting' (CROSSROADS) trial used a prospective randomized controlled design to compare the effects of changes in diet composition alone or combined with weight loss with an exercise-only control intervention on body composition and adipose tissue deposition in older adults³⁸. The trial included three arms that met weekly for the first 24 weeks of the intervention, then every 2 weeks for the remainder of the 12-month intervention. The study protocol included 30 min of group discussion related to a dietary, exercise or behavioral topic, followed by 30 min of supervised exercise using prescribed resistance-band exercises. Similarly, the 'Comprehensive Assessment of Long-Term Effects of Reducing Intake of Energy' (CALERIE) trial studied the effects of 2 years of calorie restriction on biomarkers of longevity among people who are not obese⁸⁶. Part of the CALERIE intervention included group sessions to help the participants to adhere to 25% calorie restrictions. These trials further demonstrate group dynamics.

Similar to cRCTs, IRGT trials also have nonindependence in observations that need to be accounted for during design, analysis and interpretation. Less attention has, however, been paid to the unique design and related analytic methods needed for IRGT trials. Correlations (indexed by the ICC coefficient) may develop over time in IRGT trials as group members share the treatment environment, violating the assumption that model residuals are independent within conditions. Regarding design, there is a need to account for the cluster effect. Variance inflation factors based on estimates of ICC are an important part of sample size estimation that require sample sizes to be increased compared with individual RCTs. Not accounting for this would lead to an underpowered trial. Estimating the variance inflation factor is further complicated compared with cRCTs because each arm or condition may have a different ICC coefficient. Further, the design may not have the same hierarchical structure in all conditions, which would imply a heterogeneous variance-covariance structure, allowing for ICC in the intervention condition but not in the control condition. Regarding analyses, standard linear regression assuming independence would lead to inflated type I error rates. This may prompt researchers to overestimate the significance of their findings, or to deem interventions inappropriate because they were found effective only because of statistical artifacts.

Solutions to some of these concerns can be gleaned from a simulation study. In 2018, Candlish and colleagues compared the following techniques to assess the bias, coverage and type I error: a standard

linear regression model that assumes independence; a fully clustered mixed-effects model with singleton clusters (that is, clusters containing one individual) in the control arm; a fully clustered mixed-effects model with one large cluster in the control arm; a fully clustered mixed-effects model with pseudo-clusters in the control arm; a partially nested homoscedastic mixed-effects model; and a partially nested heteroscedastic mixed-effects model⁸⁵. The simulation study found that ignoring even small ICCs results in inflated type I error rates and over-coverage of confidence intervals⁸⁵. Accounting for heteroscedasticity in mixed-effects models allowed for appropriate control of type I error rates and unbiased ICC estimates and maintained the statistical efficiency in terms of power. Wider adoption of these analytic approaches is necessary, and the simulation article provides code to implement these different variations of mixed-effect models⁸⁵. Aging-related trials such as CALERIE and CROSSROADS should in future be analyzed using mixed-effect models that account for heteroscedasticity. IRGT trials may also present scenarios where a treatment is administered to participants through multiple groups. We refer readers to simulation studies with recommendations⁸⁷. Finally, consider presenting estimates of ICC when using IRGT trials. This would help in sample size determination and design of future trials and with the interpretation of intervention group effects.

Two-stage randomized design

The assumption that one study participant's treatment assignment has no effect on another study participant breaks down in settings where study participants cannot be isolated. It is almost impossible to limit the effect of an intervention (for example, vaccines in aging populations or assisted-living interventions to reduce falls) on other group members (see additional examples in ref. 88). Interference can result in a severe understatement of treatment impacts if it is ignored. In some settings, two-stage randomized designs can address and estimate interference.

When interference is likely, two-stage randomized designs can estimate not only the average direct causal effects, but the average indirect effects (that is, interference effects), total causal effects and overall causal effects under certain assumptions. In a two-stage nested randomized design, these effects can be isolated when groups (community) are first randomized to treatments, and then at the second stage, units in the group (family) are randomly assigned at varying probabilities to the treatment levels^{6,88,89}.

For example, Halloran and Hudgens⁸⁸ consider a vaccine efficacy study whereby geographically separate groups (residential areas/clusters) are randomized to two assignment regimens (vaccine coverage). In one group, 30% of individuals are randomly assigned to receive a vaccine, and in the other, more than 50% of individuals are assigned to receive a vaccine^{6,90}. The random assignment of residential clusters to vaccine coverage represents the first stage of the two-stage randomization (for example, A or B). The second stage is done by randomly selecting who will get the vaccine in varying probabilities within the assignments at the first stage (for example, 30% of individuals are assigned to receive the vaccine in A, and 50% of individuals are assigned to receive vaccine in B). This design permits estimation of both the direct causal effect of the vaccine program (difference in disease incidence between vaccinated and unvaccinated) and, because vaccine coverage is not equal in A and B, the indirect effect of the vaccine in reducing the community spread of the infectious agent to unvaccinated individuals. The example illustrates that the vaccination effect would be underestimated when only direct effects could be estimated (that is, if all participants were randomly assigned at 50% probability). The estimation of effects from this design requires the assumptions of mixed assignment being used at each randomization stage, and stratified interference (for example, an individual's outcome from an intervention within a geriatric rehabilitation unit will be the same regardless of which other individuals receive the intervention⁶).

There are some considerations to implementing two-staged randomization under various scenarios and work is actively ongoing to address them. One such scenario is when the sizes of the randomized groups differ. In this case, the causal estimands proposed in Halloran and Hudgens may be biased. To overcome this issue, Basse and Feller proposed additional estimators for unequal group sizes⁹¹. In their example, the second stage of randomization assigns only within those units assigned to 'treatment' in the first stage; those in the control group are not randomized again. Also, the assumption of partial interference or no interference across groups holds if the groups are separated enough in both time and space. This may not occur if they share a geographical location, for example, resulting in an added complexity for the estimation of interference effect. This topic is an active area of methodologic research with potentially vast application in the analysis of complex aging research data. For more about these methodologic developments, we direct the reader to Tchetgen et al.¹

A different form of staged randomization similarly provides utility under conditions that carry expectation effects. Whereas traditional RCTs isolate the effect of treatment assignment, under 'real-world' conditions, expectations may modify the total effect. For instance, although participants can be masked to drug assignment in a trial, their prescription of the drug by a physician is not, and the expectation of knowing that a participant is not receiving a placebo may add to or subtract from outcomes. To estimate the effect of treatment assignment under 'actual conditions of use' without the use of deception, George et al. proposed 'randomization 2 randomization probabilities', whereby study participants are first randomized to a probability between 0 and 1 from a distribution defined on the unit interval⁹². Then, the participants are told their probability of being assigned a treatment (but not the actual assignment), and therefore their expectations of receiving the treatment are manipulated. To estimate expectation effects, terms are included in the statistical model for treatment assignment and probability and randomization probability-by-treatment interaction. This design is limited to treatments that can be masked from participants and entails a reduction in statistical power that needs to be considered in sample size planning.

Conclusions

Our purpose was to bring attention to the presence of interdependency in aging research studies and to present possible strategies for addressing such interdependency. Research requires tradeoffs between laboratory, clinical and real-world conditions and an understanding of ecologically valid experiments relative to the laboratory. If interdependency is suspected, investigators should account for it in the analytic model and provide proper reporting. Single-stage randomization is not always the most appropriate design, so other possible design strategies can be considered, including cRCTs (analyze as randomized), pseudo-cluster-randomized studies (enroll enough clusters guided by proper power analyses), or two-stage randomization. In addition, investigators should consider reporting ICCs for any clusters (for example, agar plates, vials, cages and housing facilities). It is easy to overlook the intersection of these issues in the clinical setting, especially because addressing them can be so challenging in a real-world setting. Every research question requires an appropriate research design; thus, interdependency does not have a single solution and may itself be the topic of interest.

References

1. Tchetgen, E. J. T. & VanderWeele, T. J. On causal inference in the presence of interference. *Stat. Methods Med. Res.* **21**, 55–75 (2012).
2. Razzoli, M. et al. Social stress shortens lifespan in mice. *Aging Cell* **17**, e12778 (2018).
3. Islam, M. et al. Effect of the resveratrol rice DJ526 on longevity. *Nutrients* **11**, 1804 (2019).

4. Manski, C. F. Identification of treatment response with social interactions. *Econom. J.* **16**, S1–S23 (2013).
5. Hong, G. & Raudenbush, S. W. Evaluating kindergarten retention policy: a case study of causal inference for multilevel observational data. *J. Am. Stat. Assoc.* **101**, 901–910 (2006).
6. Hudgens, M. G. & Halloran, M. E. Toward causal inference with interference. *J. Am. Stat. Assoc.* **103**, 832–842 (2008).
7. Kerr, J. et al. Cluster randomized controlled trial of a multilevel physical activity intervention for older adults. *Int. J. Behav. Nutr. Phys. Act.* **15**, 32 (2018).
8. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
9. National Institute on Aging. Strategic directions for research, 2020–2025. <https://www.nia.nih.gov/> (2020).
10. Lucanic, M. et al. Impact of genetic background and experimental reproducibility on identifying chemical compounds with robust longevity effects. *Nat. Commun.* **8**, 14256 (2017).
11. Bansal, A., Zhu, L. J., Yen, K. & Tissenbaum, H. A. Uncoupling lifespan and healthspan in *Caenorhabditis elegans* longevity mutants. *Proc. Natl Acad. Sci. USA* **112**, E277–E286 (2015).
12. Ayyadevara, S., Alla, R., Thaden, J. J. & Shmookler Reis, R. J. Remarkable longevity and stress resistance of nematode PI3K-null mutants. *Aging Cell* **7**, 13–22 (2008).
13. Hoffman, J. M., Dudeck, S. K., Patterson, H. K. & Austad, S. N. Sex, mating and repeatability of *Drosophila melanogaster* longevity. *R. Soc. Open Sci.* **8**, 210273 (2021).
14. Chapman, T., Liddle, L. F., Kalb, J. M., Wolfner, M. F. & Partridge, L. Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. *Nature* **373**, 241–244 (1995).
15. Prowse, N. & Partridge, L. The effects of reproduction on longevity and fertility in male *Drosophila melanogaster*. *J. Insect Physiol.* **43**, 501–512 (1997).
16. Yamamoto, R., Palmer, M., Koski, H., Curtis-Joseph, N. & Tatar, M. Aging modulated by the *Drosophila* insulin receptor through distinct structure-defined mechanisms. *Genetics* **217**, iyaa037 (2021).
17. Paigen, B. et al. Physiological effects of housing density on C57BL/6J mice over a 9-month period. *J. Anim. Sci.* **90**, 5182–5192 (2012).
18. Miller, R. A. et al. An Aging Interventions Testing Program: study design and interim report. *Aging Cell* **6**, 565–575 (2007).
19. Overton, J. M. & Williams, T. D. Behavioral and physiologic responses to caloric restriction in mice. *Physiol. Behav.* **81**, 749–754 (2004).
20. Rikke, B. A. et al. Strain variation in the response of body temperature to dietary restriction. *Mechanisms Ageing Dev.* **124**, 663–678 (2003).
21. Speakman, J. R. & Keijer, J. Not so hot: optimal housing temperatures for mice to mimic the thermal environment of humans. *Mol. Metab.* **2**, 5–9 (2012).
22. Ikeno, Y. et al. Housing density does not influence the longevity effect of calorie restriction. *J. Gerontol. A Biol. Sci. Med. Sci.* **60**, 1510–1517 (2005).
23. Smith, D. L. Jr., Yang, Y., Hu, H. H., Zhai, G. & Nagy, T. R. Measurement of interscapular brown adipose tissue of mice in differentially housed temperatures by chemical-shift-encoded water-fat MRI. *J. Magn. Reson. Imaging* **38**, 1425–1433 (2013).
24. Koisumi, A. et al. A tumor preventive effect of dietary restriction is antagonized by a high housing temperature through deprivation of torpor. *Mechanisms Ageing Dev.* **92**, 67–82 (1996).
25. Lipman, R. D., Gaillard, E. T., Harrison, D. E. & Bronson, R. T. Husbandry factors and the prevalence of age-related amyloidosis in mice. *Lab. Anim. Sci.* **43**, 439–444 (1993).
26. Nagy, T. R., Krzywanski, D., Li, J., Meleth, S. & Desmond, R. Effect of group vs. single housing on phenotypic variance in C57BL/6J mice. *Obes. Res.* **10**, 412–415 (2002).
27. Asch, S. E. In *Groups, Leadership and Men: Research in Human Relations* (ed. H. Guetzkow) 177–190 (Carnegie Press, 1951).
28. Pasupathi, M. Age differences in response to conformity pressure for emotional and nonemotional material. *Psychol. Aging* **14**, 170–174 (1999).
29. Snyder-Mackler, N. et al. Social determinants of health and survival in humans and other animals. *Science* **368**, eaax9553 (2020).
30. Epel, E. S. & Lithgow, G. J. Stress biology and aging mechanisms: toward understanding the deep connection between adaptation to stress and longevity. *J. Gerontol. A Biol. Sci. Med. Sci.* **69**, S10–S16 (2014).
31. Egami, N. Identification of causal diffusion effects under structural stationarity. Preprint at <https://doi.org/10.48550/arXiv.1810.07858> (2018).
32. Manski, C. F. Identification of endogenous social effects: the reflection problem. *Rev. Econ. Stud.* **60**, 531–542 (1993).
33. Lemaitre, M. et al. Effect of influenza vaccination of nursing home staff on mortality of residents: a cluster-randomized trial. *J. Am. Geriatrics Soc.* **57**, 1580–1586 (2009).
34. Sandvik, R. K. et al. Impact of a stepwise protocol for treating pain on pain intensity in nursing home patients with dementia: a cluster randomized trial. *Eur. J. Pain.* **18**, 1490–1500 (2014).
35. Teerenstra, S., Melis, R. J. F., Peer, P. G. M. & Borm, G. F. Pseudo cluster randomization dealt with selection bias and contamination in clinical trials. *J. Clin. Epidemiol.* **59**, 381–386 (2006).
36. Vu, T., Harris, A., Duncan, G. & Sussman, G. Cost-effectiveness of multidisciplinary wound care in nursing homes: a pseudo-randomized pragmatic cluster trial. *Fam. Pract.* **24**, 372–379 (2007).
37. Beauchamp, M. R. et al. Group-based physical activity for older adults (GOAL) randomized controlled trial: exercise adherence outcomes. *Health Psychol.* **37**, 451–461 (2018).
38. Haas, M. C. et al. Calorie restriction in overweight seniors: response of older adults to a dieting study: the CROSSROADS randomized controlled clinical trial. *J. Nutr. Gerontol. Geriatrics* **33**, 376–400 (2014).
39. Tong, G. et al. Impact of complex, partially nested clustering in a three-arm individually randomized group treatment trial: a case study with the wHOPE trial. *Clin. Trials* **19**, 3–13 (2021).
40. Fine, P., Eames, K. & Heymann, D. L. ‘Herd Immunity’: a rough guide. *Clin. Infect. Dis.* **52**, 911–916 (2011).
41. Spatola, C. A. et al. The ACTonHEART study: rationale and design of a randomized controlled clinical trial comparing a brief intervention based on Acceptance and Commitment Therapy to usual secondary prevention care of coronary heart disease. *Health Qual. Life Outcomes* **12**, 22 (2014).
42. Ogburn, E. L. Challenges to estimating contagion effects from observational data. In *Complex Spreading Phenomena in Social Systems* (eds Lehmann, S. & Ahn, Y.-Y.) 47–64 (Springer, 2017).
43. Caselli, G. et al. Family clustering in Sardinian longevity: a genealogical approach. *Exp. Gerontol.* **41**, 727–736 (2006).
44. Atzmon, G. et al. Genetic variation in human telomerase is associated with telomere length in Ashkenazi centenarians. *Proc. Natl Acad. Sci. USA* **107**, 1710–1717 (2009).
45. Rasmussen, S. H. et al. Improved cardiovascular profile in Danish centenarians? A comparative study of two birth cohorts born 20 years apart. *Eur. Geriatr. Med.* **13**, 977–986 (2022).
46. Poulain, M., Chambre, D. & Pes, G. M. Centenarians exposed to the Spanish flu in their early life better survived to COVID-19. *Aging* **13**, 21855–21865 (2021).

47. Basse, G., Ding, P., Feller, A. & Toulis, P. Randomization tests for peer effects in group formation experiments. Preprint at <https://arxiv.org/abs/1904.02308> (2019).
48. Pavela, G. et al. Packet randomized experiments for eliminating classes of confounders. *Eur. J. Clin. Invest.* **45**, 45–55 (2015).
49. Vazquez-Bare, G. Identification and estimation of spillover effects in randomized experiments. *J. Econometrics*, <https://doi.org/10.1016/j.jeconom.2021.10.014> (2022)
50. Sacerdote, B. Experimental and quasi-experimental analysis of peer effects: two steps forward. *Annu. Rev. Econ.* **6**, 253–272 (2014).
51. Gadbury, G., Coffey, C. & Allison, D. Modern statistical methods for handling missing repeated measurements in obesity trial data: beyond LOCF. *Obes. Rev.* **4**, 175–184 (2003).
52. Escoffery, C. et al. Internet use for health information among college students. *J. Am. Coll. Health* **53**, 183–188 (2005).
53. Noonan, D. & Simmons, L. A. Navigating nonessential research trials during COVID19: the push we needed for using digital technology to increase access for rural participants? *J. Rural Health*. **37**, 185–187 (2021).
54. Charness, N. & Boot, W. R. A grand challenge for psychology: reducing the age-related digital divide. *Curr. Dir. Psychol. Sci.* **31**, 187–193 (2022).
55. Newell, D. J. Intention-to-treat analysis: implications for quantitative and qualitative research. *Int. J. Epidemiol.* **21**, 837–841 (1992).
56. Taguchi, A., Wartschow, L. M. & White, M. F. Brain IRS2 signaling coordinates lifespan and nutrient homeostasis. *Science* **317**, 369–372 (2007).
57. Kernan, W. N., Viscoli, C. M., Makuch, R. W., Brass, L. M. & Horwitz, R. I. Stratified randomization for clinical trials. *J. Clin. Epidemiol.* **52**, 19–26 (1999).
58. Lachin, J. M. *Biostatistical Methods: the Assessment of Relative Risks*. Vol. 509 (John Wiley & Sons, 2009).
59. Koch, G. G., Amara, I. A., Davis, G. W. & Gillings, D. B. A review of some statistical methods for covariance analysis of categorical data. *Biometrics* **38**, 563–595 (1982).
60. Wang, R., Lagakos, S. W., Ware, J. H., Hunter, D. J. & Drazen, J. M. Statistics in medicine—reporting of subgroup analyses in clinical trials. *N. Engl. J. Med.* **357**, 2189–2194 (2007).
61. Downie, L. E. et al. Appraising the quality of systematic reviews for age-related macular degeneration interventions: a systematic review. *JAMA Ophthalmol.* **136**, 1051–1061 (2018).
62. Kalache, A. et al. Nutrition interventions for healthy ageing across the lifespan: a conference report. *Eur. J. Nutr.* **58**, 1–11 (2019).
63. Montgomery, J. M., Nyhan, B. & Torres, M. How conditioning on posttreatment variables can ruin your experiment and what to do about it. *Am. J. Political Sci.* **62**, 760–775 (2018).
64. Robins, J. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *J. Chronic Dis.* **40**, 139S–161S (1987).
65. Almirall, D., Ten Have, T. & Murphy, S. A. Structural nested mean models for assessing time-varying effect moderation. *Biometrics* **66**, 131–139 (2010).
66. Westreich, D. et al. The parametric g-formula to estimate the effect of highly active antiretroviral therapy on incident AIDS or death. *Stat. Med.* **31**, 2000–2009 (2012).
67. List, E. O. et al. The effects of weight cycling on lifespan in male C57BL/6J mice. *Int. J. Obes.* **37**, 1088–1094 (2013).
68. Murray, D. M. *Design and Analysis of Group-Randomized Trials*. Vol. 29 (Oxford University Press, 1998).
69. National Institutes of Health. Parallel Group- or Cluster-Randomized Trials (GRTs). <https://researchmethodsresources.nih.gov/methods/grt> (accessed 14 April 2021).
70. Campbell, M. K., Piaggio, G., Elbourne, D. R. & Altman, D. G. Consort 2010 statement: extension to cluster randomised trials. *BMJ* **345**, e5661 (2012).
71. Brown, A. W. et al. Best (but oft-forgotten) practices: designing, analyzing, and reporting cluster randomized controlled trials. *Am. J. Clin. Nutr.* **102**, 241–248 (2015).
72. Kimura, M. et al. Community-based intervention to improve dietary habits and promote physical activity among older adults: a cluster randomized trial. *BMC Geriatr.* **13**, 8 (2013).
73. Bolzern, J., Mnyama, N., Bosanquet, K. & Torgerson, D. J. A review of cluster randomized trials found statistical evidence of selection bias. *J. Clin. Epidemiol.* **99**, 106–112 (2018).
74. Campbell, M. K., Grimshaw, J. M. & Elbourne, D. R. Intracluster correlation coefficients in cluster randomized trials: empirical insights into how should they be reported. *BMC Med. Res. Methodol.* **4**, 9 (2004).
75. Li, F., Tian, Z., Bobb, J. & Papadogeorgou, G. Clarifying selection bias in cluster randomized trials: estimands and estimation. *Clin. Trials* **19**, 33–41 (2022).
76. Eldridge, S. M., Ashby, D. & Kerry, S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *Int. J. Epidemiol.* **35**, 1292–1300 (2006).
77. Kahan, B. C., Li, F., Copas, A. J. & Harhay, M. O. Estimands in cluster-randomized trials: choosing analyses that answer the right question. *Int. J. Epidemiol.* <https://doi.org/10.1093/ije/dyac131> (2022).
78. Hitchings, M. D. T., Lipsitch, M., Wang, R. & Bellan, S. E. Competing effects of indirect protection and clustering on the power of cluster-randomized controlled vaccine trials. *Am. J. Epidemiol.* **187**, 1763–1771 (2018).
79. Hemming, K., Taljaard, M., Moerbeek, M. & Forbes, A. Contamination: how much can an individually randomized trial tolerate. *Stat. Med.* **40**, 3329–3351 (2021).
80. Jamshidi-Naeini, Y. et al. A practical decision tree to support editorial adjudication of submitted parallel cluster randomized controlled trials. *Obesity* **30**, 565–570 (2022).
81. Borm, G. F., Melis, R. J. F., Teerenstra, S. & Peer, P. G. Pseudo cluster randomization: a treatment allocation method to minimize contamination and selection bias. *Stat. Med.* **24**, 3535–3547 (2005).
82. Melis, R. J. F., Teerenstra, S., Olde Rikkert, M. G. M. & Borm, G. F. Pseudo cluster randomization: balancing the disadvantages of cluster and individual randomization. *Eval. Health Prof.* **34**, 151–163 (2010).
83. Pence, B. W. et al. Balancing contamination and referral bias in a randomized clinical trial: an application of pseudo-cluster randomization. *Am. J. Epidemiol.* **182**, 1039–1046 (2015).
84. National Institutes of Health. Individually Randomized Group-Treatment (IRGT) Trials. <https://researchmethodsresources.nih.gov/methods/irgt> (accessed 1 July 2022).
85. Candlish, J. et al. Appropriate statistical methods for analysing partially nested randomised controlled trials with continuous outcomes: a simulation study. *BMC Med. Res. Method.* **18**, 105 (2018).
86. Ravussin, E. et al. A 2-year randomized controlled trial of human caloric restriction: feasibility and effects on predictors of healthspan and longevity. *J. Gerontol. A Biol. Sci. Med. Sci.* **70**, 1097–1104 (2015).
87. Andridge, R. R., Shoben, A. B., Muller, K. E. & Murray, D. M. Analytic methods for individually randomized group treatment trials and group-randomized trials when subjects belong to multiple groups. *Stat. Med.* **33**, 2178–2190 (2014).
88. Halloran, M. E. & Hudgens, M. G. Dependent happenings: a recent methodological review. *Curr. Epidemiol. Rep.* **3**, 297–305 (2016).

89. Philipson, T. External treatment effects and program implementation bias. NBER working paper no. T0250 <https://www.nber.org/papers/t0250> (2000).
90. Ali, M. et al. Herd immunity conferred by killed oral cholera vaccines in Bangladesh: a reanalysis. *Lancet* **366**, 44–49 (2005).
91. Basse, G. & Feller, A. Analyzing two-stage experiments in the presence of interference. *J. Am. Stat. Assoc.* **113**, 41–55 (2018).
92. George, B. J. et al. Randomization to randomization probability: estimating treatment effects under actual conditions of use. *Psychol. Methods* **23**, 337–350 (2018).
93. Chow, S.-C. & Liu, J.-p. *Design and Analysis of Clinical Trials: Concepts and Methodologies*. Vol. 507 (John Wiley & Sons, 2008).
94. Klar, N. & Donner, A. Design effects. *Wiley StatsRef: Statistics Reference Online* (2014).
95. Plewis, I. & Hurry, J. A multilevel perspective on the design and analysis of intervention studies. *Educational Res. Eval.* **4**, 13–26 (1998).
96. Bloom, H. S. Randomizing groups to evaluate place-based programs. In *Learning More from Social Experiments: Evolving Analytic Approaches* (ed. Bloom, H. S.) 115–172 (Russell Sage Foundation, 2005).
97. Shadish, W. R., Cook, T. D. & Campbell, D. T. *Experimental and Quasi-experimental Designs for Generalized Causal Inference* (Houghton Mifflin, 2002).
98. Rhoads, C. H. The implications of ‘contamination’ for experimental design in education. *J. Educ. Behav. Stat.* **36**, 76–104 (2011).
99. National Institutes of Health. Research Methods Resources: Group- or Cluster-Randomized Trials (GRTs). <https://researchmethodsresources.nih.gov/methods/grt> (accessed 1 July 2022).
100. Rubin, D. B. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Stat. Med.* **26**, 20–36 (2007).
101. Cox, D. R. *Planning of Experiments* (Wiley, 1958).
102. Neyman, J. & Iwaskiewicz, K. Statistical problems in agricultural experimentation. *Suppl. J. R. Stat. Soc.* **2**, 107–180 (1935).
103. Rubin, D. B. Randomization analysis of experimental data: the Fisher randomization test comment. *J. Am. Stat. Assoc.* **75**, 591–593 (1980).
104. Cohen, M. S. et al. Effect of bamlanivimab vs placebo on incidence of COVID-19 among residents and staff of skilled nursing and assisted living facilities: a randomized clinical trial. *JAMA* **326**, 46–55 (2021).
105. Allee, W. C. Co-operation among animals. *Am. J. Sociol.* **37**, 386–398 (1931).
106. Balthazart, J. et al. *Molecular and Cellular Basis of Social Behavior in Vertebrates*. Vol. 3 (Springer Science & Business Media, 2012).
107. Ludewig, A. H. et al. Larval crowding accelerates *C. elegans* development and reduces lifespan. *PLoS Genet.* **13**, e1006717 (2017).
108. Carey, I. M. et al. Increased risk of acute cardiovascular events after partner bereavement: a matched cohort study. *JAMA Intern. Med.* **174**, 598–605 (2014).
109. Racine, E., Troyer, J. L., Warren-Findlow, J. & McAuley, W. J. The effect of medical nutrition therapy on changes in dietary knowledge and DASH diet adherence in older adults with cardiovascular disease. *J. Nutr. Health Aging* **15**, 868–876 (2011).

Acknowledgements

We thank N. Baidwan for contributions to an early version of the paper. This work was supported in part by the National Institute on Aging (grants P30 AG050886; U24 AG056053), the Gordon and Betty Moore Foundation and the National Institute of Diabetes and Digestive and Kidney Diseases (grant P30 DK056336).

Author contributions

D.B.A. conceived the original idea. D.E.C. managed and coordinated contributions from all co-authors. All co-authors contributed to the writing and editing of the manuscript.

Competing interests

P.X. is currently an employee and shareholder of Atara Biotherapeutics at submission. D.B.A. holds equity in one company (Big Sky) and he and his institutions (Indiana University and the Indiana University Foundation) have received grants, contracts, in-kind donations and consulting fees from numerous governmental agencies, non-profit organizations and for-profit organizations including litigators and dietary supplement, food, pharmaceutical, medical device and publishing companies; however, not funded nor are directly relevant to the topic herein. All other authors declare no competing interests.

Additional information

Correspondence should be addressed to David B. Allison.

Peer review information *Nature Aging* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature America, Inc. 2022