Published in final edited form as:

J Mix Methods Res. 2017 July ; 11(3): 342–354. doi:10.1177/1558689815623685.

Attention to Individuals: Mixed Methods for N-of-1 Health Care Interventions

Peter H. Van Ness¹, Terrence E. Murphy¹, and Ather Ali¹

¹Yale University School of Medicine, New Haven, CT, USA

Abstract

In several ways, randomized controlled trials represent a high standard of rigor in clinical biomedical research. Randomized controlled trials fail, however, to yield knowledge applicable to specific individuals. This article presents a methodological rationale for a mixed methods approach to n-of-1 clinical studies that attends to the preferences and concerns of individuals while attaining high standards of qualitative and quantitative rigor. An illustrative research design involving a hypothetical music therapy intervention for a quality of life outcome is examined in some detail and a concluding argument is made that the proposed mixed methods approach is especially appropriate for early-stage research interventions intended to generate explanatory hypotheses.

Keywords

mixed methods; n-of-1 study; randomization tests; music therapy

Patients and physicians are keenly interested in knowing which course of treatment will likely provide the greatest benefit with the least harm. For a given patient, at a given time, only one treatment can be administered and its outcome measured. Ideally, one would like to compare different treatments on the same patient at the same time. This, however, is not possible: one of the treatment outcomes will necessarily be missing in such circumstances. The classic research strategy for remedying this missing data problem is to randomize individual study participants to treatment and control groups and compare mean outcomes consequent to treatment. A limitation of this solution is that while results from such a randomized controlled trial may provide generalizable knowledge, they do not provide assurance that any individual patient will benefit from the tested treatment. In recent years, medical methodologists have articulated two quite different approaches to studying the health of individuals: qualitative health research (Creswell, Klassen, Plano Clark, & Smith, 2011) and n-of-1 clinical trials (e.g., crossover studies conducted in single individuals; Duan

Reprints and permissions: sagepub.com/journalsPermissions.nav.

Corresponding Author: Peter H. Van Ness, Program on Aging, Yale University School of Medicine, 300 George Street, Suite 775 New Haven, CT 06511, USA. peter.vanness@yale.edu.

Authors' Note

The contents are solely the responsibility of the authors and do not necessarily represent the official views of NIA or NCCIH.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

et al., 2014). The objective of this article is to show how these two methodological options can be productively combined in a mixed methods approach to n-of-1 clinical interventions.

In 2011, the Office of Behavioral and Social Science Research of the National Institutes of Health published a report titled “Best Practices for Mixed Methods Research in the Health Sciences.” This publication represents an acknowledgment of mixed methods research as a valuable component of a broader national health research agenda. Among “typical reasons for using mixed methods research” listed in the report is to “add information about individuals” (Creswell et al., 2011, p. 2). Several years later, in 2014, the Agency for Healthcare Research and Quality of the Department of Health and Human Services published a report titled “Design and Implementation of N-of-1 Trials: A User’s Guide.” In the introduction to this report, the authors note the failure of randomized controlled trials to provide results applicable to individuals and document a resurgence of interest in strategies to generate individual treatment effects. Regarding the pertinence of their report topic, they write, “Of these [strategies], n-of-1 trials provide the most direct route to estimating the effect of a treatment on an individual” (Duan et al., 2014, Chap. 1, p. 1).

There are political incentives for federally funded researchers and public health practitioners to address the needs of large portions of the voting public, and there are economic pressures for corporate researchers to deliver health products and services to large markets. These incentives and pressures tend to decrease the attention that the health care system and its providers give to individual preferences and concerns. Although this relative inattention may be understandable from a practical point of view, it is nevertheless often tragic from a personal point of view. This is especially evident in the case of rare diseases that have not been extensively investigated by universities and corporations, despite government encouragement as expressed in the Orphan Drug Act of 1983 (Wellman-Labadie & Zhou, 2010). Among the more frequently cited reasons for interest in complementary and alternative medicine (CAM) is the perception that the dominant health care system poorly attends to individual patient needs. The National Center for Complementary and Integrative Health (formerly, the National Center for Complementary and Alternative Medicine), established in its current form in 1999, is a governmental acknowledgment of the public’s increasing interest in CAM (Briggs & Killen, 2013). A more recent federal initiative to encourage health care researchers to identify “which approaches to care might work best, given their [patients’] particular concerns, circumstances, and preferences” (PCORI, 2014) is the establishment of the Patient-Centered Outcome Research Institute in 2010 (Clancy & Collins, 2010).

The recent methodological interest in mixed methods research and n-of-1 clinical trials occurs, therefore, in a contemporary intellectual context of efforts to integrate epidemiologic research of human populations with clinical case research applicable to specific individuals. Furthermore, for the first author of this article, this methodological interest has specifically emerged from the American context of pragmatist philosophy. In the essay, “What Pragmatism Means” (1879/1963a), William James wrote, “The whole function of philosophy ought to be to find out what definite difference it will make to you and me, at definite instants of our life, if this world-formula or that world-formula be the true one” (p. 25). In “The Will to Believe” (1896/1963b), then writing principally about religious beliefs

instead of philosophical ideas, James (1896/1963b) described an individualistic orientation that is equally applicable to most people's assessment of proposed treatment regimens: "Deadness and aliveness in hypotheses are not intrinsic properties, but relations to the individual thinker. They are measured by his (or her) willingness to act" (p. 194). Appreciation of practical differences to individuals is a recommendation for mixed methods research and n-of-1 clinical trials. James expressed frequent disinterest, and occasional distrust, in logic, mathematics, and statistics for their capacity to lure researchers astray from grappling with the practical concerns of individuals. John Dewey's (1938/1991) *Logic: The Theory of Inquiry* similarly shuns symbolic formalisms and faults them when they are conceived apart from "human beings in the ordinary or 'natural' processes of living" (p. 26). This pragmatist valuation by James and Dewey of individual interests and goals in belief formation and their insistence that beliefs be engaged in a self-corrective process of reasoning informed by practical action provide strong motivations for complementing quantitative methods of n-of-1 studies with qualitative inquiry into the preferences and concerns of individuals.

Charles Sanders Peirce, the first pragmatist and an accomplished logician and mathematician, represents a somewhat different sensibility, one that sought to wed pragmatic meaning with logical and mathematical rigor. Indeed, in "Pragmatism as the Logic of Abduction" (1903) he used the phrase "logic of abduction" to describe the systematic cultivation of explanatory hypotheses, and he sought thereby to chart a path from surprising aspects of encounters with individual phenomena to the formulation of testable hypotheses whose confirmations yield scientific laws that then become the surest guidelines for habitual practices promoting human well-being (Peirce, 1903/1998). Although Dewey's pragmatism has been positively cited by qualitative researchers since the seminal work of Glaser and Strauss (1967), Peirce's thought has been less utilized. Peirce's variety of pragmatism, though, provides the philosophical basis for this effort to integrate mixed methods with n-of-1 trial designs in a way that will both attend to individuals' interest in treatments beneficial for themselves and to the social desideratum of research programs leading to replicable and generalizable knowledge.

The methodological approach to n-of-1 studies recommended in this article involves several stages of qualitative and quantitative inquiry. These stages make up a sequence that concludes with a qualitative reflection about individuals intended to achieve what the French phenomenological philosopher Paul Ricoeur (1967) called "a second naiveté" (p. 351). In this perspective, the role of quantitative analysis of n-of-1 studies in a mixed methods approach is not to settle issues in an abstract way but rather to sensitize researchers to descriptive patterns of individual experience that might not otherwise be fully appreciated.

Music Therapy: Individual Benefits From Personalized Interventions

Randomized controlled trials balance prognostic factors in treatment groups, avoid selection bias in treatment assignments, and provide a basis for statistical inference by randomly assigning study participants to treatment and control groups (Duan, Kravitz, & Schmid, 2013). They attain their reputation for rigor by employing additional study design elements such as the blinding of study participants and researchers to the particular treatment groups

to which participants are assigned. N-of-1 clinical trials employ randomization and they sometimes employ blinding (Duan et al., 2013). For instance, in such trials, individual osteoarthritis patients may be randomly assigned to periods taking acetaminophen or ibuprofen for pain relief. The capsules for these medications can be made similar so that the patient is unaware of what medication he or she is taking at a given period of time. For behavioral interventions, such blinding is often not possible (Boutron, Tubach, Giraudeau, & Ravaud, 2004). Yet given an n-of-1 trial's orientation toward individual effectiveness rather than causal efficacy, blinding is often not of paramount importance (Duan et al., 2014). Blinding helps avoid placebo effects and other psychological mechanisms of action that can be confused with a medication's causal efficacy. Patients, however, are most concerned with an achieved benefit and less concerned with how the benefit is obtained. For this reason, n-of-1 behavioral interventions are most illustrative of this study design's orientation toward individual benefit. Of course, not all behavioral interventions are suitable for n-of-1 studies. N-of-1 studies generally require that treatments address chronic rather than acute conditions, that treatment effects be stable and quick in onset rather than episodic and slow in onset, and that carryover effects be minimal (Lillie et al., 2011).

Music therapy interventions are prime examples of unblinded behavioral interventions. They are also illustrative of the individual orientation of qualitative methods because a basic finding of this area of research is that people respond more positively to music that has personal meaning for them, for example, music associated with positive experiences or fond memories or that continues religious or ethnic traditions (Clair & Memmott, 2008). The first stage in the sequence of research proposed in this article is to interview people in an open-ended qualitative fashion to determine their musical preferences. Music therapy has increasingly attracted the interest of mixed methods researchers; the recent review article by Bradt, Burns, and Creswell (2013) ably summarizes relevant methodological approaches and recent work in this area.

Located within the larger domain of CAM (National Center for Complementary and Alternative Medicine, 2013), music therapy involves mind-body interventions in which interactions between the mind, body, and behavior are used to affect health-related outcomes. Music therapy interventions are often conducted with small samples and sometimes are reported as descriptive case studies (Engström & Hammar, 2012; Raglio, Bellandi, Baiardi, Gianotti, & Granieri, 2012). In many instances, these choices are attributable to limited funding, but they may also reflect the fact that such interventions often yield results that are highly individual and may vary by context. Small sample sizes and orientations toward individual factors make music therapy a fitting context for including both quantitative and qualitative components, that is, mixed methods (Bradt et al., 2013). The qualitative part of the study usually involves questions to which study participants provide verbal responses in their own words such as responses about musical preferences (Carr et al., 2012; Chlan et al., 2013; Chlan & Heiderscheit, 2009). A combination of small sample sizes and mixed methods might connote a lack of methodological rigor to some researchers. Studies, though, can be designed so that randomization tests are properly employed in concert with a qualitative component that includes questions regarding key assumptions for sound inference. Such designs will contribute to a mixed methods approach

to music therapy and other behavioral interventions that can provide informative results that enhance and extend traditional case study methods.

Weak Study Designs as Lost Opportunities

Research in music and medicine has advanced considerably in recent years. A recent issue of *Music and Medicine* was devoted to the topic of “Neurobiological Aspects of Music and Music Therapy” (Stegemann, 2013). In the articles in this issue, a diverse group of researchers from North America, Europe, and Asia identified possible neurobiological mechanisms that might account for positive results of behavioral music therapy interventions. Such new research represents one variety of response to the 2005 plan instituted by the American Music Therapy Association (2005) to promote the evidence-based practice of music therapy in a way similar to earlier, more general efforts toward evidence-based medicine (Sackett, Straus, Richardson, Rosenberg, & Haynes, 2000).

Concerns about the methodological rigor of some music therapy studies remain evident in the research literature (Zarit & Orrell, 2012), especially in studies having small samples. One response to such concerns is to refrain from inferential claims by reporting purely descriptive results and no inferential results (Engström & Hammar, 2012; Raglio et al., 2012). This approach risks lost opportunities.

For instance, in the recent pilot study by Engström and Hammar (2012), a humming intervention was administered at mealtimes to two older persons having dementia with the objectives of making caregiver feeding more effective and making patient eating more enjoyable. Small sample interventions with quality of life outcomes are especially pertinent (Sole, Mercadai-Brotons, Gallego, & Riera, 2010) but also especially challenging (Van Ness et al., 2010) for research involving older persons. This humming intervention study outcome was observed in a sophisticated manner using video observations that were subsequently converted into measurements on the Edinburgh Feeding Evaluation (EdFED; Stockdell & Amelia, 2008). This scale has an overall range from 0 to 20, with lower scores indicating fewer feeding/eating problems. Study results were reported for two preintervention baseline measurements, two intervention measurements, and one postintervention follow-up measurement observed over the course of 5 weeks; two participants demonstrated mean reductions of 5.5 and 3.5 points during the intervention on the EdFED. No inferential statistics were calculated and small sample size was noted among its limitations.

The particular circumstances of any given intervention may constrain the types of study designs it can employ. With modest modifications, however, some music therapy interventions can be conducted with a considerably more informative study design than a purely descriptive case report. Including an element of randomization in a study design is commonly regarded as a methodological strength. With a sample size of two, the traditional uses of randomization, being the random sampling of experimental units from a well-defined population and the random allocation of intervention protocols to experimental units, are not feasible. The remaining alternative is to introduce a random timing of intervention initiation. This is a strategy commonly used in n-of-1 studies and other small sample experiments (Dugard, File, & Todman, 2012). Unlike case studies, which can be simply descriptive, n-

of-1 studies are intrinsically interventional, longitudinal, and inferential, even though inference does not characteristically extend beyond the limited sample to some larger population of interest (Auerbach & Zeitlin, 2014).

Peirce is widely credited with being among the first to incorporate randomized elements in experimental study designs (Peirce & Jastrow, 1885). Although the randomized component he introduced to a psychological study was not the same as is used in n-of-1 crossover designs, his willingness to innovate in this way is characteristic of a pragmatist impulse to draw on a plurality of methods to solve practical research problems. The combination of quantitative n-of-1 methods with qualitative methods in a mixed methods study design represents this same pragmatist spirit when conducting research that tries to satisfactorily attend to individual preferences and concerns.

A brief description of a simple hypothetical n-of-1 study will be given in which intervention protocol and outcome measurement are modeled on the Engström and Hammar (2012) case study, but whose study design incorporates the random timing of intervention initiation. The described designed modifications will provide an opportunity for drawing inferences about intervention effectiveness at reducing feeding/eating problems and for showing how the assumptions and meanings of such inferences can be explicated by qualitative study components.

A Methodological Thought Experiment

A caregiver feeding an older person with moderate dementia is videotaped for 30 noon mealtimes over the course of a calendar month, some without, and then some with, a humming mealtime intervention. A minimum of five mealtimes is prespecified as occurring initially without and, finally, with the intervention so that the number of mealtimes at which the continuous days (or phase) of intervention might be randomly initiated is 20 (see Figure 1, for a graphical presentation of the study design). The null hypothesis for this experiment is that there will be no difference between mean EdFED scores for the days before and after initiating the humming intervention.

The test is one-sided with the expectation that intervention scores will be lower than preintervention scores, indicating fewer feeding/eating problems during the intervention phase. Tests results will yield a probability value calculated by dividing the number of test statistics equal to or more extreme (relative to the null hypothesis) than the test statistic hypothetically observed by the total number of test statistics obtained from calculations yielded by all possible days for the initiation of the intervention (Edgington & Onghena, 2007). This “*p* value” is interpretable as the observed significance level or, more specifically, assuming the truth of the null hypothesis, as the probability of observing a value of the test statistic that is at least as extreme as the one hypothetically observed and computed from sample data.

Table 1 provides a hypothetical data set for an intervention experiment following this study design. The intervention column of data shows that the day randomly chosen (by software) to initiate the intervention was Day 13. Table 2 provides three calculated test statistics: those

corresponding to this hypothetically observed experiment and for two of the 19 alternative days for the initiation of the intervention, one with a longer and another with a shorter intervention phase. The test statistic for the hypothetically observed experiment, that is, the difference in means resulting from the subtraction of intervention scores from preintervention scores, is larger than the two alternatives. Thus, the hypothetically observed experiment result yielded smaller mean EdFED values for the intervention phase than for the preintervention phase, indicating fewer feeding problems. Table 1 shows that all of the EdFED scores in the intervention phase are equal to or smaller than scores for the nonintervention phase. This implies that no test statistic for the preintervention phase that might have been observed if the intervention had been initiated on some other possible intervention day is more extreme than the test statistic hypothetically observed. The p value for this hypothetical experiment is $1/20$ or $.05$ because only 1 day out of the 20 possible days for intervention initiation yielded a test statistic that was equal to or more extreme than the one hypothetically observed. This estimation of a probability value constitutes the second, quantitative stage in the design sequence for a mixed methods approach to n-of-1 studies.

A Matter of Interpretation

Including an intervention in a small sample experiment and estimating a p value for its results are designed to provide evidence for whether the person or persons in the study sample benefited from the intervention. Small sample studies are often exploratory, intending to yield insights for future study rather than grounds for decisive action. For this reason, it is more appropriate to interpret the p value as a Fisherian test of significance rather than as a Neyman–Pearson hypothesis test (Christensen, 2005). Accordingly, the test has a null hypothesis, that is, that there is no difference in mean scores for the preintervention and intervention phases, but no explicit alternative hypothesis, and the notion of a rejection region is inapplicable. A Fisherian significance test is a probabilistic indirect argument whose p value reflects the weight of evidence for rejecting the null hypothesis (Fisher, 1960/1990). The weight of evidence approach to interpreting statistical tests has tended toward Bayesian interpretations that emphasize the continual updating of an inference in light of new and different types of evidence rather than all or nothing rejections of null hypotheses. (It is again worth noting that Peirce was among the first to use the phrase “weight of evidence” in the interpretation of statistical tests, although he did not develop his thinking on this topic in what is now understood as a Bayesian or subjective view of probability; on this point, he appears to be inconsistent, at least in his early writings, with the general themes of his pragmatist philosophy [Peirce, 1878/1992].) In this context, probability is more appropriately understood to mean degree of belief than long-run relative frequency because the reference set, that is, the total number of logically possible outcomes of a randomization process, is not a sampling distribution in the sense of imagined repetitions of the observed experiment.

Properly interpreting the study null hypothesis and its corresponding p value is only the first step in understanding the sort of inference yielded by a small sample intervention study. The inferences made from them have almost no external validity or generalizability even though these studies may be useful for hypothesis generation or for understanding anecdotal reports. With so few participants, such study results are neither representative of a study population

nor characteristic of a specific stochastic process. This second characteristic is what requires a nonparametric approach to inference such as a randomization test; the first characteristic requires that inference apply only to the study sample and not to some population from which it was drawn. Test precision may be increased somewhat, for instance, by expanding the reference set to include a final nonintervention phase in which the intervention is withdrawn, yielding a nonintervention–intervention–nonintervention study design. Precision may also be increased by including more than one participant in the study, each with their own randomly selected intervention initiation date; Bayesian methods have recently been proposed for combining information from similar n-of-1 studies (Duan et al., 2013). Such methods will make possible smaller p values, but in most circumstances, it is doubtful that they will yield appreciably greater external validity.

Assessing Assumptions

An inference is robust to the extent that it remains valid under different assumptions. The best small sample intervention studies cannot generally be as robust as well-conducted large sample studies: In this design context, even small changes will likely have a large impact. Nevertheless it is crucial that researchers carefully assess key study assumptions and modify their degree of belief in study results in light of how well such assumptions are satisfied. Rubin (2004) identifies the “Stable Unit Treatment Value Assumption (SUTVA)” as the key set of assumptions for experimental studies (p. I-2.1). These assumptions include that there be only one version of the intervention for the treatment group and only one form for the control group, and that the randomization status for one experimental unit have no impact on the outcome for any other. The first SUTVA requirement can be satisfied in a well-conducted small sample study, although when music therapy interventions are personalized in the sense of allowing study participants to choose their own favored music, this assumption deserves special scrutiny (Chlan et al., 2013). The second requirement presents major challenges for n-of-1 studies and related crossover designs. Of course, upon choice of relevant inferential test statistic, such as a t test, other assumptions become pertinent and should be assessed.

The SUTVA requirement was originally applied to experiments in which individual experimental units are randomly allocated to intervention and control groups. In the illustrative n-of-1 experiment described above, observation occurrences are construed to be the randomized experimental units. Thus, SUTVA applies to the randomization status of an observation at one time point having no impact on the outcome at some second time point. This requirement prohibits carryover effects from one intervention phase to a subsequent phase. The use of n-of-1 study designs for quality of life music therapy interventions more plausibly meets this requirement than some other intervention modalities because the impact of any given administration of the intervention is unlikely to be long-lasting.

Other SUTVA violations are also possible in n-of-1 studies. For example, during the nonintervention phase, there can be a trend in study outcomes over time attributable to participant changes unrelated to the intervention. In cases like the music therapy study described above, it is unlikely, though possible, that an individual’s cognitive status would increase or decline in a short period of time. In the context of an experiment involving very

different experimental units but some design similarities, Rubin suggested that an alternative test statistic might help address suspected SUTVA violations. Instead of calculating a difference of means for the nonintervention and intervention phases, Rubin (2004) suggested subtracting the first postintervention score value of a member of the reference set from its immediate preintervention predecessor. Informally, if the intervention were rapidly effective, one would expect the difference between immediate pre–post scores to be among the largest calculated for the reference set. This test statistic could be expected to lessen the impact of secular trends in the nonintervention group and cumulative intervention effects in the intervention group, though it provides no guarantee of a lack of bias due to a carryover SUTVA violation.

An Application of Mixed Methods

A small sample intervention study involving a quality of life outcome is a natural setting for qualitative research. Qualitative methods are especially appropriate for generating rich data about little-known phenomena such as experiential CAM intervention effects (Bradley, Curry, & Devers, 2007). Ways of conducting a qualitative study may emerge from an explicit theoretical orientation, or, as in the grounded theory approach, they may lack motivation by any preconceived theory (Strauss & Corbin, 1998). Both types of qualitative interviewing are anticipated in the concluding qualitative stage of the proposed mixed methods approach to n-of-1 studies. For instance, representing a definite theoretical orientation toward statistical inference, interview questions might be asked about aspects of the intervention experience that pertain to whether key experimental assumptions, such as SUTVA, were satisfied. Study participants might be asked about the extent to which the impact of one intervention phase carried over to a subsequent phase. They might also be asked about the extent to which the intervention impact got increasingly larger as the intervention phase transpired as a way of determining whether the magnitude of the intervention was related to the duration of its application. Answers to these qualitative questions about experimental assumptions should then be considered in interpreting the *p* value from the quantitative study. Specifically, they should inform the degree of belief that the researcher gives to assessing the weight of evidence against the null hypothesis that there was no difference in mean outcomes for the preintervention and intervention phases of the music therapy intervention experiment. These qualitative questions about methodological assumptions constitute a first moment in the final qualitative stage of the mixed methods sequential design recommended here for research providing attention to individuals.

A second moment in the final qualitative stage of the design sequence of the proposed mixed methods study design would then follow. It should proceed with fewer preconceptions regarding responses. After the completion of the hypothesized intervention, study participants may be asked open-ended questions regarding the quality of their experience of eating while the caregiver feeding them was humming or singing. They may be asked about how well their musical preferences were satisfied and related qualitative questions. At this point, researchers are encouraged to adopt an interpretive stance that is not only pragmatic but also phenomenological. On analyzing quantitative data from study participants, the researchers should return to descriptions of individually lived experiences without reducing them to a level of discourse that the study participants would not recognize; they should seek

to attain a variety of Ricoeur's (1967) "second naiveté." Doing so is pragmatic in the sense of acknowledging that different methodological techniques are often useful for different study purposes.

Some Controversial Issues

The mixed methods literature is rife with methodological controversy. One view looks with suspicion on mixed methods approaches whose "motive is to validate quantitative statistical findings with qualitative data results" (Hesse-Biber, 2010, p. 457). In this context, it has been suggested that casting qualitative research in a supportive role, as a "handmaiden" to quantitative research (Hesse-Biber, 2010), is expressive of a commitment to "positivism" as a methodological and epistemological stance. This is certainly not the approach advocated here even though the qualitative moment inquiring about study assumptions on completion of the quantitative n-of-1 study is intended to promote a better interpretation of n-of-1 quantitative results. The interpretation of randomization tests as Fisherian significance tests whose probabilistic interpretation, and whose integration with qualitative results, rests on the idea of degree of belief is incompatible with positivist approaches to probability theory (Carnap, 1962; von Mises, 1951/1981). The approach adopted here is consistent with an editorial in the *Journal of Mixed Methods Research*: "From a pragmatic perspective, the primary issue is to determine what data and analyses are needed to meet the goals of the research and answer the questions at hand" (Bazeley, 2009, p. 203). Furthermore, a pragmatist attention to individuals, such as illustrated by quotations given above from William James, is a prime rationale for the study designs proposed here. The question at issue in the hypothetical study described here is whether a music therapy intervention improves the quality of the feeding and eating experience for an older person with some degree of dementia. The use of randomization in a small sample intervention and the use of significance testing for assessing study results are intended to allow inference as to whether the intervention was effective for at least one study participant. Qualitative methods can provide additional insight regarding the interpretation of this inference using traditional strategies for quantitative/qualitative integration (O'Cathain, Murphy, & Nicholl, 2010). Also, when addressed to experimental assumptions, they can help assess the validity of inference to members of the sample in a way that complements rather than obstructs quantitative methods (Bryman, 2007). Equally, though, quantitative methods are often enlisted to enrich a qualitative understanding of some individual phenomenon. In this hypothetical illustration, a quantitative n-of-1 study is used to assess the overall impact on individuals of a music therapy intervention designed to improve the experience of eating.

Conclusion

A major theme of this report has argued that n-of-1 studies can be especially helpful in attending to the preferences and concerns of individuals when combined with qualitative components in a mixed methods approach. Presupposed by this argument is the belief that randomization tests can make important contributions to the analysis of n-of-1 studies. Implied by this argument is the conviction that by attending to individual preferences and concerns, the mixed methods approach to n-of-1 studies is especially important in early-

stage clinical research. In concluding this article, some summary reflections about this presupposition and this implication are given.

Randomization tests are not the only statistical tool for analyzing small sample studies (Aickin, 2004; Shih, Ohman-Strickland, & Lin, 2004), though they do have the advantage of relative simplicity (Rubin, 1980). Also, as presented here using an example of a music therapy intervention, they allow for an intuitive way to integrate qualitative and quantitative results. Randomization tests are useable for pre- and postintervention phase studies, but they are also applicable to more complicated n-of-1 crossover designs, for instance, in which multiple pairs of crossover phases are used, and to small experiments in which experimental units are randomized to intervention and control groups. They can also be used for outcomes other than quality of life outcomes; for instance, they can be used in assessing safety procedures and quality of care improvements that are important to the health of individuals. They have an advantage over other nonparametric methods, such as an exact version of the Wilcoxon rank-sum test, because the test statistic for the latter is not clinically meaningful, that is, a sum of ranks, while the randomization test is flexible as to the nature of the test statistic used in calculations of elements in the study reference set. Hence, because they are relatively simple in application, clinically meaningful in interpretation, and readily integrated with qualitative results, randomization tests are methodological tools that researchers should consider when conducting mixed methods small sample health care interventions.

Attention to individual preferences and concerns is especially important in early-stage research. Writing in the *Journal of Complementary and Alternative Medicine*, Aickin (2007) states that “the true purpose of an early phase study is to guide the expenditure of future research effort, to decide which paths should be followed and which should be abandoned” (p. 447). For this purpose, failures can be as informative as successes. They contribute to the social process of trial and error by which beliefs are updated and improved in a community of inquiry. Dewey (1991) in his logical writings especially emphasized that logical thinking is not so much a formal derivation from first principles but an iterative encounter with experience that succeeds through a social process of trial and error by which sound habits of inference are established. As Aickin notes, it is better to conduct unsuccessful studies with small samples than with large ones. The experience of individuals as represented in results of combined qualitative and n-of-1 studies can make valuable contributions to early-stage research by advancing Peirce’s process of abduction in which promising explanatory hypotheses are generated.

Acknowledgments

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: P30AG021342—Claude D. Pepper Older Americans Independence Center at the Yale University School of Medicine—Dr. Van Ness and Dr. Murphy; and K23AT006703—National Center for Complementary and Integrative Health at the National Institutes of Health—Dr. Ali.

References

- Aickin M. Separation tests for early-phase complementary and alternative medicine comparative trials. *Evidence-Based Integrative Medicine*. 2004; 1:225–231.
- Aickin M. The importance of early phase research. *Journal of Alternative and Complementary Medicine*. 2007; 13:447–450. [PubMed: 17532739]
- American Music Therapy Association. Strategic priority on research: Overview. 2005. Retrieved from http://www.musictherapy.org/research/strategic_priority_on_research/overview
- Auerbach, C., Zeitlin, W. *SSD for R*. New York, NY: Oxford University Press; 2014.
- Bazeley P. Integrating data analyses in mixed methods research. *Journal of Mixed Methods Research*. 2009; 3(3):203–207.
- Boutron I, Tubach F, Giraudeau B, Ravaud P. Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. *Journal of Clinical Epidemiology*. 2004; 42:1758–1772.
- Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: Developing taxonomy, themes, and theory. *Health Services Research*. 2007; 42:1758–1772. [PubMed: 17286625]
- Bradt J, Burns DS, Creswell JW. Mixed methods research in music therapy research. *Journal of Music Therapy*. 2013; 50:123–148. [PubMed: 24156190]
- Briggs JP, Killen J. Perspectives on complementary and alternative medicine research. *Journal of the American Medical Association*. 2013; 310:691–692. [PubMed: 23990064]
- Bryman A. Barriers to integrating quantitative and qualitative research. *Journal of Mixed Methods Research*. 2007; 1(1):8–22.
- Carnap, R. *Logical foundations of probability*. 2nd. Chicago, IL: University of Chicago Press; 1962.
- Carr C, d'Ardenne P, Sloboda A, Scott C, Wang D, Priebe S. Group music therapy for patients with persistent post-traumatic stress disorder: An exploratory randomised controlled trial with mixed methods evaluation. *Psychology and Psychotherapy: Theory, Research and Practice*. 2012; 85:179–202.
- Chlan LL, Heiderscheit A. A tool for music preference assessment with critically ill patients receiving mechanical ventilatory support. *Music Therapy Perspectives*. 2009; 27:42–47. [PubMed: 24489432]
- Chlan LL, Weinert CR, Heiderscheit A, Tracy MF, Skaar DL, Guttormson JL, Savik K. Effects of patient-directed music intervention on anxiety and sedative exposure in critically ill patients receiving mechanical ventilatory support: A randomized clinical trial. *Journal of the American Medical Association*. 2013; 309:2335–2344. [PubMed: 23689789]
- Christensen R. Testing Fisher, Neyman, Pearson, and Bayes. *The American Statistician*. 2005; 59:121–126.
- Clair, AA., Memmott, J. *Therapeutic uses of music with older adults*. Silver Spring MD: American Music Therapy Association; 2008.
- Clancy C, Collins FS. Patient-Centered Outcomes Research Institute: The intersection of science and health care. *Science Translational Medicine*. 2010; 2 37cm18.
- Creswell, JW. *Qualitative inquiry and research design: Choosing among five approaches*. 3rd. Thousand Oaks, CA: Sage; 2013.
- Creswell, JW., Klassen, AC., Plano Clark, VL., Smith, KC. Best practices for mixed methods research in the health sciences (for the Office of Behavioral and Social Sciences Research). 2011. Retrieved from http://obssr.od.nih.gov/mixed_methods_research/
- Dewey, J. *Logic: The theory of inquiry*. New York, NY: Holt, Rinehart & Winston; 1991. Original work published 1938
- Duan, N., Eslick, I., Gabler, NB., Kaplan, HC., Kravitz, RL., Larson, EB., Vohra, S. Kravitz, RL., Duan, N., editors. *Design and implementation of n-of-1 trials: A user's guide*. 2014. Retrieved from <http://effectivehealthcare.ahrq.gov/ehc/products/534/1844/n-1-trials-report-130213.pdf>

- Duan N, Kravitz RL, Schmid CH. Single patient (n-of-1) trials: A pragmatic clinical decision methodology for patient-centered comparative effectiveness research. *Journal of Clinical Epidemiology*. 2013; 66:S1–S28. [PubMed: 23849143]
- Dugard, P., File, P., Todman, J. Single-case and small-n experimental designs: A practical guide to randomization tests. 2nd. New York, NY: Routledge; 2012.
- Edgington, ES., Onghena, P. Randomization tests. 4th. Boca Raton, FL: Chapman & Hall/CRC; 2007.
- Engström G, Hammar LM. Humming as a potential tool for facilitating feeding situations between persons with dementia and their caregiver: A single case study. *Music and Medicine*. 2012; 4:231–236.
- Fisher, RA. Statistical methods, experimental design, and statistical inference: A re-issue of statistical methods for research workers, the design of experiments, and statistical methods and scientific inference. Oxford, England: Oxford University Press; 1990. Design of experiments; p. 11-26. Original work published 1960
- Glaser, BG., Strauss, AL. The discovery of grounded theory: Strategies for qualitative research. Chicago, IL: Aldine; 1967.
- Hesse-Biber S. Qualitative approaches to mixed methods research. *Qualitative Inquiry*. 2010; 16:455–468.
- James, W. Pragmatism and other essays. New York, NY: Washington Square Press; 1963a. Original work published 1879
- James, W. The will to believe. In: James, W., editor. Pragmatism and other essays. New York, NY: Washington Square Press; 1963b. p. 193-213. Original work published 1896
- Lillie EO, Patay B, Diamant J, Issell B, Topol EJ, Schork NJ. The n-of-1 clinical trial: The ultimate strategy for individualizing medicine. *Personalized Medicine*. 2011; 8:161–173. [PubMed: 21695041]
- National Center for Complementary and Alternative Medicine. Complementary, alternative, or integrative health: What's in a name?. 2013. NCCAM Publication No. D347 Retrieved from <http://nccam.nih.gov/health/whatisccam>
- O'Cathain A, Murphy E, Nicholl J. Three techniques for integrating data in mixed methods studies. *British Medical Journal*. 2010; 341:c4587. [PubMed: 20851841]
- PCORI. Why PCORI was created. 2014. Retrieved from <http://www.pcori.org/content/why-pcori-was-created>
- Peirce, CS. The probability of induction. In: Houser, N., Kloesel, C., editors. The essential Peirce: Selected philosophical writings. Vol. 1. Bloomington: Indiana University Press; 1992. p. 155-169. Original work published 1878
- Peirce, CS. Pragmatism as the logic of abduction. In: Houser, N., Kloesel, C., editors. The essential Peirce: Selected philosophical writings. Vol. 2. Bloomington: Indiana University Press; 1998. p. 226-241. Original work published 1903
- Peirce CS, Jastrow J. On small differences in sensation. *Memoirs of the National Academy of Sciences*. 1885; 3:73–83.
- Raglio A, Bellandi D, Baiardi P, Gianotti M, Granieri E. Music therapy in frontal temporal dementia: A case report. *Journal of the American Geriatrics Society*. 2012; 60:1578–1579. [PubMed: 22889022]
- Ricoeur, P. The symbolism of evil. New York, NY: Harper & Row; 1967.
- Rubin DB. Randomization analysis of experimental data: The Fisher randomization test (a commentary). *Journal of the American Statistical Association*. 1980; 75:591–593.
- Rubin, DB. Basic concepts of statistical inference for causal effects in experiments and observational studies. 2004. Retrieved from http://s3.amazonaws.com/zanran_storage/www.s-a-e.org.ar/ContentPages/2492390951.pdf
- Sackett, DL., Straus, SE., Richardson, WS., Rosenberg, W., Haynes, RB. Evidence-based medicine: How to practice and teach EBM. New York, NY: Churchill Livingstone; 2000.
- Shih WJ, Ohman-Strickland PA, Lin Y. Analysis of pilot and early phase studies with small samples. *Statistics in Medicine*. 2004; 23:1827–1842. [PubMed: 15195318]

- Sole C, Mercadai-Brotons M, Gallego S, Riera M. Contributions of music to aging adults' quality of life. *Journal of Music Therapy*. 2010; 12:264–281.
- Stegemann T. Neurobiological aspects of music and music therapy. *Music and Medicine*. 2013; 5:202–247.
- Stockdell R, Amelia EJ. The Edinburgh Feeding Evaluation in Dementia Scale: Determining how much help people with dementia need at mealtime. *American Journal of Nursing*. 2008; 108:46–54. [PubMed: 18664760]
- Strauss, AL., Corbin, JM. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. 2nd. Thousand Oaks, CA: Sage; 1998.
- Van Ness PH, Charpentier PA, Ip EH, Leng I, Murphy TE, Tooze JA, Allore HG. Gerontologic biostatistics: The statistical challenges of clinical research with older study participants. *Journal of the American Geriatrics Society*. 2010; 58:1386–1392. [PubMed: 20533963]
- von Mises, R. *Probability, statistics and truth*. New York, NY: Dover; 1981. Original work published 1951
- Wellman-Labadie O, Zhou Y. The US Orphan Drug Act: Rare disease research stimulator or commercial opportunity? *Health Policy*. 2010; 95:216–228. [PubMed: 20036435]
- Zarit SH, Orrell M. Editors' introduction to the addendum to efficacy of music therapy based on cycles of sessions: A randomized controlled trial. *Aging & Mental Health*. 2012; 16:264. [PubMed: 22224755]

Figure 1.
Diagrammatic representation of the design of the intervention study with one study participant.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1
Edinburgh Feeding Evaluation Scores (Hypothetical) for a Music Intervention Study.

Day	Intervention	Score	Day	Intervention	Score
1	0	15	16	1	9
2	0	16	17	1	10
3	0	17	18	1	8
4	0	14	19	1	7
5	0	12	20	1	6
6	0	13	21	1	7
7	0	15	21	1	8
8	0	16	23	1	10
9	0	15	24	1	11
10	0	14	25	1	11
11	0	13	26	1	11
12	0	12	27	1	10
13	1	12	28	1	9
14	1	8	29	1	10
15	1	7	30	1	8

Table 2

Illustrative Calculation of Test Statistics for Edinburgh Feeding Evaluation Scores for a Randomization Test Reference Set.

Day of intervention commencement	Preintervention scores (Y_A)	Intervention scores (Y_B)	Test statistic ($\bar{Y}_A - \bar{Y}_B$)
Day 7 (possibly earlier)	15, 16, 17, 14, 12, 13	15, 16, 15, 14, 13, 12, 12, 8, 7, 9, 10, 8, 7, 6, 7, 8, 10, 11, 11, 11, 10, 9, 10, 8	$(\bar{Y}_A)=14.500$, $(\bar{Y}_B)=10.292$, $(\bar{Y}_A - \bar{Y}_B)=4.208$
Day 13 (hypothetically observed)	15, 16, 17, 14, 12, 13, 15, 16, 15, 14, 13, 12	12, 8, 7, 9, 10, 8, 7, 6, 7, 8, 10, 11, 11, 11, 10, 9, 10, 8	$(\bar{Y}_A)=14.333$, $(\bar{Y}_B)=9.000$, $(\bar{Y}_A - \bar{Y}_B)=5.333$
Day 17 (possibly later)	15, 16, 17, 14, 12, 13, 15, 16, 15, 14, 13, 12, 12, 8, 7, 9	10, 8, 7, 6, 7, 8, 10, 11, 11, 11, 10, 9, 10, 8	$(\bar{Y}_A)=13.000$, $(\bar{Y}_B)=9.000$, $(\bar{Y}_A - \bar{Y}_B)=4.000$