Developing a Systematic Framework for Utilizing Discrete Types of Qualitative Data as Therapy Research Evidence

ARTHUR C. BOHART a, c, MAKENNA BERRY b, & CARRIE WICKS b

^a Saybrook University, San Francisco, CA, and California State University—Dominguez Hills, Carson, CA

Email: abohart@saybrook.edu

ABSTRACT

In a follow-up study on a "Research Jury Method" for qualitatively adjudicating statements about psychotherapy process and outcome (Bohart, Tallman, Byock, & Mackrill, 2011), the utility of 56 criteria for aiding jurors in analyzing a case history was examined. It was found that these criteria were useful in helping jurors identify aspects of the case that led to their conclusions. Utilizing the criteria also helped jurors adjudicate differences between them.

Key words: Research Jury Method, adjudicated case study method, case law, psychotherapy process and outcome, case-specific changes, case studies, clinical case studies

INTRODUCTION

In a previous study (Bohart, Tallman, Byock, & Mackrill, 2011) we described a "Research Jury Method" for qualitatively adjudicating statements about psychotherapy process and outcome. The current study is a further exploration of that method. Based on a suggestion by Daniel Fishman (personal communication, 2009), it focuses on the usefulness of utilizing specific criteria to help qualitatively parse the evidence in a case study.

The adjudicational method is a single case method. It was originally developed (e.g., Bohart & Humphreys, 2000) as an alternative to the randomized controlled trial model utilized in empirically supported treatments (ESTs) research (Task Force, 1995), because we believed it fit the nature of humanistic psychotherapy better. The adjudicational method is based on the metaphor of a jury trial. In a trial, jurors have a whole case record to use. They look for patterns in the evidence that lead them to make judgments that the client is guilty or not in a criminal trial, or liable for damages in a civil trial.

As applied to psychotherapy, the idea is to present individuals who act as jurors with a sufficiently rich case record and to ask them to comb it for evidence in order to answer two basic questions: (a) did the client change, and (b) if so, is there evidence that psychotherapy caused or contributed to the change? This is a qualitative-hermeneutic method. In one sense the method

^b Saybrook University, San Francisco, CA

^c Correspondence concerning this article should be addressed to Arthur C. Bohart, Saybrook University, 747 Front St. 3rd Floor, San Francisco, CA 94111-1920.

contrasts with other kinds of qualitative research where the researcher goes in with more of a pure phenomenological orientation. The idea in such research is to start out with some sort of provisional coding scheme and let it develop in order to see what is there. An example would be grounded theory (Strauss & Corbin, 1997). In the adjudicational method jurors go in with clear questions that focus them more specifically on evidence relevant to the questions they want to answer. The method is interpretive in that jurors are not merely coding this or that bit of data but looking for whole patterns which plausibly support one conclusion or the other. Nonetheless it is still discovery-oriented. Jurors in trials are supposed to suspend judgment as much as possible until they see all the evidence, and then let the evidence teach them. Similarly, in the case of our study, jurors go in to see if there is sufficient evidence to build an argument that the client has changed, or to decide there is no sufficient evidence to do that. Then, if they have sufficient evidence that the client has changed, additionally to see if there is evidence to build an argument that therapy contributed to the change process.

Bohart, Tallman, Byock, and Mackrill (2011) have previously discussed why they believe this method falls within the realm of science. The logic of how conclusions are drawn is different than in a randomized controlled trial. It does not depend on strict control but rather upon drawing conclusions based upon a convergence of evidence. Conclusions are based on looking for a pattern that builds to a plausible conclusion. Plausibility is ultimately the criterion rather than a statistical decision rule. David Rennie (1995) has talked about plausibility being a criterion in qualitative research. He has also pointed out that this is actually part of quantitative science as well, but we cannot go into that here.

That is the basic logic. In one study done using this method six graduate students listened to tapes of 6-8 session therapies with an open mind, took notes, and made judgments as to whether clients changed or not and whether therapy helped (Bohart & Humphreys, 2000; Humphreys, Bohart, & Dutile, 2000). A second study (Bohart, Tallman, et al., 2011) utilized a detailed case study written up by Thomas Mackrill (2008; reprinted in this issue as Mackrill, 2011). It consisted of 39 single spaced pages of the case of Jane, working with her psychodynamic/existential therapist Joe for 12 sessions. The case history included extensive quotes from client and therapist diaries as well as from tapes of the 12 therapy sessions. Three jurors converged in agreeing that the client had changed moderately and that therapy had helped.

In both studies jurors were instructed to go through the case material and make decisions on their own evaluation of the evidence. However, they were given a set of 56 criteria (see Appendix 1) to use if they so desired to help them parse the evidence. They were asked to read over the criteria before reading the case. They were told that they did not need to use the criteria unless they wanted to. Thus the criteria were not presented as a formal rating system or interpretive framework, but rather as a set of heuristics.

These criteria had been developed in two ways. One was through logical analysis and the other was through discussions with the original research team based on listening to pilot tapes. The 56 criteria were developed to help jurors draw plausible conclusions. In part they were designed to help deal with some of the issues that quantitative designs attempt to control for.

This included designing them to help jurors plausibly conclude that alternative hypotheses were not likely.

The criteria were broken up into three groups. The first group consists of heuristics to help decide if the client has changed or not (see items 1-11 in Appendix 1). For instance, one is that the client says that he or she has changed. However, clients could say this to please the researcher. Or it could be a kind of halo effect. So in addition other criteria help minimize these possibilities. For instance, if the client can supply supporting detail it makes it more plausible that the change is indeed real. This supporting detail might be revealed in things the client says during sessions. If the client reports things that did not change, or if the client reports different degrees of change in his or her presenting problems this also helps rule out the possibility that any given change is merely due to a halo effect or a desire to please the researcher.

The second group deals with specific changes (see items 12-39 in Appendix 1). Again, we look for supporting evidence. If a client says he or she is no longer depressed is there supporting evidence in the form of how the client acts in therapy. Does the client report other things consistent with being less depressed—such as taking more initiative, trying new things out, sleeping better, and so on.

The third group has to do with evidence that therapy contributed to the changes observed (see items 40-56 in appendix 1). One bit of evidence might be that the client says that therapy contributed, but there has to be plausible evidence supporting this, such as that the client reports detail on how therapy helped, or the changes are plausibly related to what went on in therapy. In addition, in any therapy of more than a few sessions, the client will report details of his or her life in the therapy session. The juror has to be able to rule out that life events did not plausibly account for the change.

It should be reiterated that the criteria were not developed to be used as formal rules of evidence for judgment. If they had been, it might be possible to turn them into formal criteria, such as those used by Staudinger, Smith, and Baltes (1994) in their studies on wisdom. In those studies judges have five general criteria they use to decide if someone shows wisdom or not. The five criteria are broken down into more specific sub-criteria. Raters make interpretations of complex interviews, not unlike the kinds of interpretations made by our jurors. However, after making such interpretations, they ultimately assign numbers to how much wisdom is shown. Raters are trained to use the criteria until they both agree with one another (reliability) and agree with pre-established criteria (validity).

It is conceivable that one could do something like that with our criteria. However, that is not how juries work. Juries depend, in some sense, on the opposite of training to reliability. Although juries are given rules of evidence, it is up to them to apply them in their own individual way. What is wanted is *naturally occurring reliability* rather than trained reliability. What one wants in a jury are people coming from diverse perspectives, seeing the evidence from diverse points of view, perhaps focusing on different aspects of the evidence, yet all converging to the same conclusion. It is precisely this that lends truth value to the conclusion, in contrast to the idea of training everyone to see things in the same way. Bohart, Hewitt, Heilmann, & Threlkeld

(1976) once used a similar method. One half of a group of psychologically naive students was randomly assigned to receive interpersonal training, and the other half, no training. After the students in the first group had received the training, the students in both groups were interviewed by other students who rated them on empathy and other interpersonal qualities. No attempt was made to train the student interviewers to a criterion of reliability. Bohart et al. did not want that. They wanted the student interviewers to react to the trained students in their naturally occurring ways. If, with this diversity built in, the trained students were rated higher after training in interpersonal skills than the untrained students, it lent confidence that Bohart et al. were indeed providing the trained students with skills that might be useful in the real world.

For a similar reason we did not want our jurors to follow formal rules for applying the criteria. We wanted them to apply them, combine them, and interpret them in their own diverse ways in order to make their judgments. Thus, while they were criteria that could help "capture" and identify qualitative evidence that could be used to make decisions, they were still utilized in qualitative, hermeneutic ways, rather than in ways based in quantitative, positivistic thinking.

The current study is an attempt to evaluate the potential of these criteria for use in a more systematic way for capturing qualitative data.

METHOD

Two graduate students were enlisted to go through the case of Jane (Mackrill, 2008, 2011), the same case studied by Bohart, Tallman, Byock, and Mackrill (2011). The two graduate students both had some experience in doing counseling, although neither was a licensed counselor. They first followed the same procedure used in the previous study. They read over the list of 56 criteria, but then went through the case on their own. They were told they could use the criteria or not, but to primarily to rely on their own judgment. They were to make the same two judgments made in previous work: (a) did Jane change, and (b) if so, did therapy contribute? After they made these judgments (which made this study in part a replication of Bohart et al., this issue) they were asked to go back through the case and, using the 56 criteria, (a) decide for each criterion of it applied to the case or not, and (b) if they believed that it did, cite some evidence in support of that. This was followed by having the two jurors meet to discuss their differences on criteria where they disagreed. They were instructed to discuss their differences but told they did not have to move towards a consensus. What was more important was to understand their reasons for differing.

RESULTS

First, as a form of replication of the previous study, the two new jurors came to similar conclusions as the three in the previous study that Jane had changed a moderate amount. They also concluded that therapy had contributed.

Second, when asked to evaluate whether each of the 56 criteria fit the case or not, the jurors used slightly different methods of deciding. One tended to use "yes" and "maybe" a lot, as well as "no" or "no evidence," while the other tended to say "yes," "no" or "no evidence."

Considering "yes" and "maybe" as agreeing, the two jurors agreed on 9 of the 11 items of the first set of criteria dealing with whether the client changed or not; and on 19 of the 26 items on the second set of criteria concerned with specific changes. They also agreed on 16 of the 17 criteria on whether therapy helped or not. One juror did not make judgments on two criteria so overall they agreed on 44 of the 54 criteria, and disagreed on 10.

Third, of the 10 criteria on which they disagreed to any extent, the two jurors discussed 8 criteria where they clearly disagreed. After discussion, the two jurors agreed on 7 of these 8.

Fourth, based on the reports from the two jurors, the reasons for their disagreements occurred for two reasons: Either one juror had noticed evidence in the case the other juror had missed, or one juror had construed or interpreted the evidence somewhat differently in a way that the other now found convincing. When these were clarified the two jurors came to agreement in 7 of 8 cases. It is interesting to note that in a study of psychology journal reviewers, Fiske and Fogg (1990) found something similar. When different reviewers disagreed it often was because they focused on different aspects of an article. This was more so than that they had literally disagreed on specific points.

Fifth, in terms of the citation of evidence, the two jurors typically gave page numbers and line numbers or paragraph numbers, or cited some bit of text without particularly explaining how the evidence fit their conclusion. Three examples follow:

Criterion 2: They mention things that make it clear that clients either did something or experienced something different than what they normally do or experience in the course of their everyday lives.

Yes, page 30, lines 26-30: Where she realizes that she can just be herself and a man can find her.

Criterion 9: They mentioned problems that didn't change.

Yes, page 27, lines 19-23: The sad/scary memories don't go away.

Criterion 23: Getting a new perspective which brings greater coherence, reduces debilitating guilt, gives new positive behavioral options, helps the person let go of something from the past.

Yes, page 18, lines4-9:

"On this date I was very conscious of standing back and giving him space to take the lead, e.g., I let him take care buying tickets, ordering sandwiches...When there were breaks in the conversation I was very conscious of not asking about him right away, which is usually what I do, almost instinctively. I tried to tell myself, that the ("embarrassing") break didn't mean anything and that I was completely relaxed about it (even though I was completely conscious of the fact that neither of us was saying anything). And then I noticed, that after a while he actually asked about me, or my work, or something else, which was a great experience. So it also went according to plan."

It would be interesting in a future study to have the jurors write out their reasons for why a given piece of evidence was seen as supporting a particular conclusion. We did not have them do that in this study.

Sixth, it was found that even when the jurors agreed on a criterion, only sometimes did they cite the same evidence in support of their decisions.

Seventh, the two jurors reported that going through the criteria was useful. While it did not change the conclusions they had previously come to, they found it helpful to more specifically identify the evidence that led to their conclusions. This facilitated more detailed conversations if there was disagreement.

DISCUSSION

The use of specific criteria for helping researchers identify specific types of qualitative data appears to be promising. The two jurors found it useful in helping them clarify their decisions. They particularly found that the sequence employed was useful: first go over the case as jurors, sifting the evidence as a juror would, and then follow that with employing criteria to test their conclusions more specifically.

The value of having jurors specifically identify what bits of data were used to support or buttress conclusions is first, that it makes the process by which they came to their conclusions potentially more transparent to anyone who might want to scrutinize how conclusions were reached in a given case. Second, it could lead to developing a detailed map of just how bits of evidence relate to final judgments of outcome. This could inform future work in this area.

It should be reiterated that the criteria were utilized in a qualitative fashion. Criteria were not applied in a mechanical fashion. They were tools used by the jurors to sift through and interpret evidence. It was up to each juror to interpret the criterion as they would, allowing for the open, discovery-oriented quality that qualitative research attempts to have.

In conclusion, identifying specific criteria for sifting evidence in a qualitative jury method can help sharpen how conclusions are drawn.

REFERENCES

- Bohart, A.C., Tallman, K.L., Byock, G. & Mackrill, T. (2011). The "Research Jury" Method: The application of the jury trial model to evaluating the validity of descriptive and causal statements about psychotherapy process and outcome. *Pragmatic Case Studies in Psychotherapy*, 7(1), Article 8, 101-144, Available: http://hdl.rutgers.edu/1782.1/pcsp_journal
- Bohart, A. C., & Humphreys, C. (2000, June). A qualitative "adjudicational" model for Assessing psychotherapy outcome. Paper presented at the meeting of the International Society for Psychotherapy Research, Chicago, Illinois.
- Bohart, A., Hewitt, B., Heilmann, A., & Threlkeld, D. (1976). Effects of paraprofessional

- training on one aspect of generalized interpersonal effectiveness. *American Journal of Community Psychology*, *4*, 309-312.
- Fiske, D. W., & Fogg, L. F. (1990). But the reviewers are making different criticisms of my paper: Diversity and uniqueness in reviewer comments. *American Psychologist*, 45(5), 591-598.
- Humphreys, C., Bohart, A., & Dutile, R. (2000, August). *Adjudication model for psychotherapy research: Psychotherapy outcome.* Paper presented at the American Psychological Association Convention, Washington, DC.
- Mackrill, T. (2008). The therapy journal project: A cross-contextual qualitative diary study of psychotherapy with adult children of alcoholics. Unpublished Doctoral Dissertation. Copenhagen University, Denmark.
- Mackrill, T. (2011). The case of "Jane and Joe": A diary-based, cross-contextual case study. *Pragmatic Case Studies in Psychotherapy*, 7(1), Article 11, 187-229. Available: http://hdl.rutgers.edu/1782.1/pcsp_journal
- Rennie, D. L. (1995). On the rhetorics of social science: Let's not conflate human science and natural science. *The Humanistic Psychologist*, 23, 321-332.
- Staudinger, U. M., Smith, J., & Baltes, P. B. (1994). *Manual for the assessment of wisdom-related performance*. Berlin: Max Planck Institute For Human Development.
- Strauss, A. C., & Corbin, J. M. (1997). *Grounded theory in practice*. Thousand Oaks, CA: Sage.
- Task Force on Promotion and Dissemination of Psychological Procedures, Division of Clinical Psychology of the American Psychological Association (1995). Training and dissemination of empirically-validated psychological treatments: Report and recommendations. *The Clinical Psychologist*, 48, 3-23.

APPENDIX 1: CRITERION LIST

EVIDENCE THAT THE PERSON CHANGED

- 1. Clients note themselves that they have changed
- 2. They mention things that make it clear that clients either did something or experienced something different than what they normally do or experience in the course of their everyday lives
- 3. Clients are relatively specific about how they have changed
- 4. They provide supporting detail
- 5. They show changes in behavior in the therapy session plausibly related to the kinds of changes they should be making outside of the session.
- 6. Plausible reports by the clients that others have noted that the client has changed
- 7. Plausible indicators reported by the client: better grades, promotion at work, less use of medication, new activities such as jogging
- 8. They mentioned problems that didn't change
- 9. They mention problems that didn't change.
- 10. The changes mentioned seem plausible given the degree of difficulty of the problem, degree of time in therapy
- 11. If there is a major change reported, it is described in rich enough detail to be plausible

SPECIFIC CHANGES

- 12. If the client comes in depressed they show a reasonably consistent change in mood; more ups than downs as therapy goes on--i.e. come to therapy less often depressed, seem less depressed, recover more quickly
- 13. If they report being anxious, they report either managing it better, or reductions in anxiety in key situations, and this shows a positive trend over therapy
- 14. If they report being unable to leave their house (agoraphobia) they report an example suggesting that they made a new and more concerted effort to go out and it met with at least some degree of success, and their affect about trying it is positive and hopeful (i.e. there is an increase in perceived possibility for them that they can do it)
- 15. If their problem is a habit problem (studying, overeating, drinking, smoking, etc.) they report concrete changes. With a habit problem ONE incident of change is not usually enough to say that a substantial change has occurred. We would want evidence that this one change was something new, or a new attempt after having been discouraged. But we would like it better if the person could report several successes; a pattern of success. But if a few fresh changes were made and the person seemed optimistic, that we could take as preliminary evidence of change
- 16. If the problem is a demoralization problem ("I can't"), or involves demoralization, the person begins to show hope and optimism--a sense of possibility, a sense that it will be a challenge. They become challenge oriented. If they fail they focus more on learning from the challenge than on what it means about them in terms of their inadequacy. In fact, they focus more on the difficulty of the task than on their inadequacies. In other words, when

they fail they no longer see it as a complete sign of their inadequacy, or their failure. If they choose not to pursue it any further it is after a reasonable evaluation where they conclude reasonably that a shift in priorities is in order, or action plan.

- 17. Evidence of new-found confidence in judgment.
- 18. Evidence of greater competence in judgment--as the individual thinks out the problem he or she does it more proactively, considers alternatives, weighs them, uses good intuition.

 Does not seem driven by fear and jump to conclusions. They weight options aloud, think things out.
- 19. Evidence of greater proactive determination and persistence in relation to a reasonable goal.
- 20. If they make a risky choice, they seem to make it in a reasonable way.
- 21. Arriving at a major decision that the person was struggling with.
- 22. Coming up with a whole new plan which is innovative.
- 23. Getting a new perspective which brings greater coherence, reduces debilitating guilt, gives new positive behavioral options, helps the person let go of something from the past
- 24. Gaining a new perspective where they seem to be acceptantly criticizing themselves, seeing their own limitations, but not in a defensive or overly critical way.
- 25. Gaining a perspective that "I am not my problem"
- 26. Identity work: clarifies fundamental goals and values. If no goals or values, begins to confront these issues. If has adopted goals and values from parents but is beginning to question them, begins to evaluate for self. If is in an "identity crisis," or moratorium, struggles with issues and makes progress in making commitments. Identity work can take place in any or all of the following areas: vocational goals, moral values, goals about relationships, goals about children, religious values, political values, values about what makes for a meaningful life, gender issues, sexuality, ethnicity and cultural background
- 27. Identity work: Real self controversies--what is my real self, am I being untrue to my real self? Movement towards some kind of reconciliation or decision.
- 28. Traumatic experiences--signs of letting go of it, coming to terms with it, reductions in symptoms such as flashbacks or nightmares, or at least a greater sense that these can be handled and not so debilitating
- 29. Achievement of specific goals--becoming more assertive, as evidence by self-report of concrete instances, perhaps seeming more assertive in the therapy session, rise in confidence
- 30. Interpersonal changes--reported changes in a positive fashion in relationships--handling anger better, less dependence, greater problem solving, greater realistic acceptance of others (i.e., but NOT accepting certain things such as abuse), greater empathy as demonstrated towards others and towards the therapist (more careful listening, less confrontive). With therapist acts more proactively, dialogically, less dependent, less aggressive, less need for dominance.
- 31. Specific changes: finished a project, made attempts to protect daughter, exercising. Made a new friend. Got and kept a job
- 32. Greater realization that there may be some things that will take ongoing work
- 33. Changes in self-relationship. Greater realization and appreciation of accomplishments; more specific and concrete and accurate assessment of talents and effort; less global, negative self-attributions; greater self-empathy; greater self-listening to intuitions, felt experiencing; greater receptive internal dialogue; holding constructs more tentatively to

- evaluate them; more of an open, searching mentality; if overinflated self-esteem or self-confidence, taking a more careful look at how one might be doing, offending people, etc.
- 34. Reduction in any presenting symptoms, such as feeling weak, fearful, tiring quickly, feeling no interest in things, feeling stressed, blaming oneself, feeling suicidal, unfulfilling sex life, feeling lonely, frequent arguments, difficulty concentrating, feeling hopeless about the future, having disturbing thoughts come to mind, upset stomach, sweating, dizziness, heart pounding, trouble getting along with others, trouble sleeping, headaches.
- 35. Increases in positive things: self-efficacy, enjoying spare time, feeling loved and wanted, greater happiness, greater sense of direction or optimism, greater acceptance of the injustices of life in a productive way.
- 36. Better ability to define goals in a proactive and functional way.
- 37. Prosocial changes--volunteering, involvement in productive activities, new projects.
- 38. Changes in physiology--less sweating, more calm and relaxed in therapy.
- 39. Changes in appearance in a positive fashion (if observed).

EVIDENCE THAT IT WAS THERAPY THAT HELPED

- 40. Clients themselves report that therapy helped.
- 41. Clients are relatively specific about how therapy helped, and it is described in a plausible way.
- 42. Outcomes are relatively specific and idiosyncratic to each client and vary from client to client (if comparing across clients).
- 43. In their reports, clients are discriminating about how much therapy helped, i.e. they do not in general give unabashedly positive testimonials.
- 44. They describe plausible links to the therapy experience.
- 45. To the rater a plausible narrative case can be made linking therapy work to positive changes. This includes the following (#46-56):
- 46. Therapy provides a work space where clients have an opportunity to talk, think, express. The things the client talks about are the things that change, or if other things change, the client notes a relationship of them to the therapy experience. Client notes that this helped.
- 47. Therapist's empathic understanding, warmth, acceptance, seems to relate to client's increased engagement, willingness to try new things, productive exploration.
- 48. Therapist's encouragement, support, positive attitude seem to be related to client's overcoming demoralization, willingness to confront challenges, not be discouraged by failure. Therapist supports client productively when client fails. Keeps eye focused on productive behavior and this seems to relate to client's doing so also.
- 49. Therapist's warmth, empathic listening, seems to provide safe atmosphere for client to confront painful experiences, and these in turn change.
- 50. Therapist's in-tune questions, reflections, interpretations, or comments, seem to facilitate clients' exploration, gaining new perspectives, developing action plans, creativity. Client feels recognized.
- 51. Clients engage in concrete procedures in therapy and changes are congruent with what they are trying to achieve, and there is evidence of these changes. Examples: EMDR--clients work through a traumatic experience and then seem relieved afterwards, and at the next session; clients engage in chair work and either resolve an internal conflict, or come to

- terms with someone they have unresolved feelings towards; and this change persists or at least partially persists in subsequent sessions; clients challenge dysfunctional cognitions and show plausible changes in mood or behavior,
- 52. Issues client struggles with in therapy change plausibly over time in accord with the trajectory of the client's working on them. E.g. client talks about them week after week, and has ups and downs, but gradually masters them, and the mastery seems related to their ongoing struggle with it in therapy. In other words, perhaps each week they talk about experiences related to resolving the problem, works on it, and gradually masters it.
- 53. Clients report changes in trajectory from their past life in the problem. Clients report something new in regard to coping with the problem, and relate it to therapy, or it seems related to therapy. Clients report a history of failed coping with the problem, and now it is changing. Even if client reports having tried some of these things before, now reports that therapy has helped have confidence in the effort and helps him or her persist.
- 54. There are no plausible life changes that could have assumed major responsibility for the change. Or, if there is a life change, it seems to be a result of therapist deliberative activity, or it gets incorporated into the therapy activity in a productive way
- 55. Topics not dealt with in therapy did not change, or, if they did change, there was a plausible reason why they changed from the therapy or from clearly independent reasons. In other words, they can be accounted for so that we can assume we are not talking about a global halo effect.
- 56. Clients' mastery experiences, problem actuation, and clarification and gaining of new perspectives that occurs in therapy are related to the changes.