Published in final edited form as:

Subst Use Misuse. 2017 September 19; 52(11): 1429–1438. doi:10.1080/10826084.2017.1284236.

Saving My Life: Dynamics of Peer and Staff Corrections Among Therapeutic Community Residents

Nathan J. Doogan and

The Ohio State University College of Public Health

Keith L. Warren

The Ohio State University College of Social Work

Abstract

Background—Therapeutic communities (TCs) maintain order and encourage behavior change through a system of peer corrections. This study is the first quantitative analysis of the dynamics of the exchange of peer corrections at TCs.

Objectives—We applied longitudinal social network analysis to compare the reactions of TC residents to peer versus staff intervention, while analyzing dynamics of correction exchange among residents.

Method—The data consisted of a large database of staff and peer affirmations and corrections at four therapeutic community units that occurred between the years 2006 and 2008. We modeled the data as a directed temporal social event network, using a generalized linear mixed effects model to analyze predictors of corrections among residents.

Results—Residents were more likely to send a correction following peer affirmations and corrections than following staff affirmations and corrections. Residents reciprocated corrections to individual peers. Autocorrelation was evident in both sending and receiving corrections and residents were more likely to send a correction after having sent an affirmation. Residents who arrived at roughly the same time were more likely to exchange corrections. Residents tended to send and receive more corrections in the middle three months of their treatment. European American residents and those with higher scores on the LSI-R were more likely to receive corrections than others.

Conclusions—TC residents respond more strongly and more positively to peer than to staff intervention. The pattern of exchange of peer corrections in TCs is complex. This suggests possible paths to improved outcomes.

Keywords

Therapeutic communities; Cooperation; Evolutionary psychology; Social network analysis; Mutual aid; Longitudinal networks; Unit atmosphere; Substance abuse treatment; Systems science

Introduction

Therapeutic communities (TCs) use a system of peer feedback as a primary clinical tool; residents are expected to correct peers for actions that contravene TC norms of “right living” as well as affirming peers for prosocial actions (De Leon, 2000; Perfas, 2012). This system of mutual monitoring and peer correction assumes that residents who are recovering from substance abuse, many of whom have committed criminal offenses, can effectively hold each other accountable for behaviors contravening TC norms. These norms include personal responsibility, honesty with oneself and the community, avoiding the mere mention of positive experiences associated with a criminal lifestyle as well as rules against theft, undue personal conflict and substance abuse (De Leon, 2000; Hawkins & Wacker, 1986; Perfas, 2012).

A willingness to correct peers is seen as a sign of responsible concern, preparing them for a world outside the TC in which they will be held accountable for their behaviors (De Leon, 2000; Hawkins & Wacker, 1986; Pearce & Pickard, 2012; Perfas, 2012). A willingness to accept a correction is seen as a sign of honesty, itself a prerequisite for any significant change, and an opportunity to connect surface behavior to underlying attitudes (Perfas, 2012). At one TC which the first author has visited, it is common for the residents to respond to a correction with the phrase, “Saving my life.” The peer who gave the correction will then reply, “Life worth saving.” Nevertheless, there is evidence that residents understandably find peer corrections to be difficult to accept (Hawkins & Wacker, 1986; Patenaude, 2005; Perfas, 2012).

No other clinical intervention places such emphasis on mutual accountability supported by a system of peer corrections. The peer-based Alcoholics Anonymous program includes a rule against criticism in meetings, although peers may offer negative feedback once the meeting is over (Makela et al, 1996). Oxford Houses, which like TCs serve as residential centers for individuals in recovery from substance abuse, have no set rules other than that residents are not allowed to drink or use drugs and will be evicted if doing so (Oxford House Manual, 2015). The only parallel we have been able to find to the strict TC system of mutual accountability is from a different field entirely; grassroots community organizations in the tradition of the Industrial Areas Foundation place a strong emphasis on maintaining accountability through mutual critiques among organizers and community leaders (Kleidman, 2004).

While clinical researchers often disparage the use of negative feedback (Andrews & Bonta, 2010), theory and empirical evidence drawn from experimental studies in evolutionary psychology, anthropology, economics and evolutionary biology support the use of mild sanctions in the maintenance of cooperation in large groups such as TCs (Barr, 2001; Bowles & Gintis, 2011; Fehr & Gächter, 2002). This literature finds that mild sanctions often promote cooperation. In several experiments simply allowing peers to criticize non-cooperators without any possibility of further sanctions dramatically increased cooperation in groups—that is, altered behavior in a more prosocial direction (Barr, 2001; Masclet, Noussair, Tucker & Villeval, 2003). This is very similar to the use of corrections in TCs (De Leon, 2000).

The difference in perspectives may stem from the implicit assumption on the part of clinical researchers that feedback comes from clinicians rather than peers; TC clinical theory and evolutionary theory assume exactly the opposite. Moreover, while evolutionary theory emphasizes maintenance of cooperation rather than individual outcomes, these two factors are linked in TCs because the atmosphere of the unit and resident attitudes toward peers predict outcomes (Carr & Ball, 2014; Mandell, Edelin, Wenzel, Dahl & Ebener, 2008; Pearce & Pickard, 2012; Ronel, Elisha, Timor & Chen, 2013; Tiburcio & Kressel, 2011). Some qualitative studies suggest that TC residents value peer feedback (Stevens, 2013), while others suggest that certain conditions can lead to resentment (Patenaude, 2005). At this point there has been no quantitative comparison of resident reactions to corrections when delivered by peers versus staff.

An analysis of the dynamics of the exchange of corrections between TC residents, and particularly a comparison of residents' responses following peer versus staff intervention, would be of considerable value in verifying TC clinical theory and understanding and improving TC clinical process. Such an analysis would face two technical challenges. First, it is necessary to track the exchange of corrections between residents over a period of time. This is not as difficult as one would think; it is common for TCs to keep records of the corrections that residents exchange in order to track clinical progress and activity in the TC. Second, since the exchange of corrections between any two individuals is not independent, it is necessary to use a statistical framework that accounts for confounding dependencies in the data. In this study, we treat the exchange of corrections as a directed social network—when A corrects B an arrow goes from A to B. We then use a social network time series model to analyze the corrections that residents send and receive as a function of previous affirmations and corrections coming from peers or from staff.

Drawing from De Leon's (2000; see also Perfas, 2012) model of the community itself as the method of treatment as well as the extensive literature on generalized reciprocity (Isen & Levin, 1972; Herne, Lappalainen & Kestila-Kekkonen, 2013; Molm, Collett & Schafer, 2007), we hypothesize that residents who have recently received an affirmation or correction from a peer will increase their own engagement with the community by becoming more likely to give corrections to peers compared with those that have recently received an affirmation or correction from staff.

Methods

Data

Data were drawn from archival records kept at four community based correctional TC units housing male residents at three TCs. Two of the TCs, one with a 90-bed male unit and the other an entirely male facility with 60-beds, drew on predominantly rural catchment areas. The third TC included two 80-bed male units and drew on a catchment area that was a mix of rural and urban counties.

Both affirmations and corrections were recorded in three steps. First, the resident who sent the affirmation/correction would write it on a form including the names of the sender and receiver, the date, and the content. Second, senior residents and staff would review the

affirmation or correction. Finally, the affirmation or correction would be read aloud to the community and then entered into a database. The result was an electronic record of daily affirmations and corrections.

These facilities also kept data on ethnicity, age, entry and departure date, and residents' scores on the Level of Services Inventory-Revised (LSI-R), a standardized measure of risk of criminal recidivism (Andrews & Bonta, 1995). The LSI-R includes subscales on substance abuse, education, previous offenses, employment, financial status, social support, family/marital status, living accommodations, recreational skills, mental health issues and attitudes toward criminal behavior. Ethnicity was recoded into a bivariate dummy variable with minority status as the reference value.

Sampling—The data sets of affirmations and corrections varied in size. The two rural TCs gathered this data for 8 years, while the two units at the mixed catchment TC gathered it for 3 years. For consistency in time frame, we randomly sampled 100 days from each TC unit within the time period from the beginning of 2006 to September of 2008. The number of days sampled and the size of the TC units meant a huge number of possible opportunities for the exchange of affirmations and corrections; the number of records needed to represent all possibilities was approximately 2,064,000. But far fewer affirmations and corrections were actually exchanged during these days (Affirmations: 12,104; Corrections: 6,582). King and Zeng (2001) recommend that no fewer than five percent of records represent events to avoid biases in parameter estimation. We therefore took a subsample as recommended by King and Zeng (2001), retaining all events and randomly sampling from all non-events such that in our final sample 20% of the data records represented a dyad (including a sender and a receiver) in which a correction occurred.

Analysis

The analysis quantified the probability of a correction event within a dyad given exposure to a complex web of similar events that occurred among the community (e.g., recently received staff versus resident affirmations or corrections). The fitted model delineated the contribution of various network related exposure types to the probability of new event occurrence. The unit of analysis was a directed dyad with a sender, a receiver, their characteristics and exposures to previous affirmations and corrections. The outcome was the presence or absence of a correction from the sender to the receiver during a given day.

To capture the complex dependencies in the data, we adapted the approach of de Nooy, (2012) and Lerner, Bussmann and Snijders (2013). We built a generalized linear mixed effects logistic regression model that accounted for multiple forms of dependency, including dependency on previous interactions.

We included a random effect to capture variation across residents in their propensity to send corrections. Similarly, we included a random effect to capture variation across residents in their propensity to *receive* corrections. Finally, many corrections may occur within a single day and there is heterogeneity across days; we therefore included a third random effect to capture variation across days.

We accounted for other forms of dependency by incorporating predictors that characterized localized exposure of residents to previous affirmations and corrections. As a simple example, a previous affirmation from my peer might increase the odds that I will reciprocally affirm that same peer. This is a form of dependency that is likely to be present in the data, that must be controlled for, and that we want to quantify the strength of. We discuss this in more detail later.

We use a flexible Markov chain Monte Carlo sampler called Stan (Stan Development Team, 2014) to obtain samples from the posterior distribution of the parameters. Stan is a Bayesian modeling framework that allows the incorporation of *a priori* knowledge (a.k.a., a “prior”) into the model parameter estimates. We admitted fairly poor *a priori* knowledge and used zero centered and diffuse priors for all fixed mean parameters (normal distribution with standard deviation of five). We used a uniform prior for the random effect variances (Gelman, 2006). Inferences are drawn from the posterior distribution of the model parameters (Gelman, Carlin, Stern & Rubin, 2004), which are represented as means and 95% credible intervals (95% CI). An estimate is deemed statistically significant if the 95% CI excludes zero. We later present hypotheses that require the comparison of model parameters, for which we report a *p*-value that represents the probability that one parameter value does not exceed another¹.

Measures

Dependent variables—The dependent variable, y_{ijt} , represents whether a new correction was given by actor i to actor j on day t . It is coded one if a new event occurred and zero otherwise.

Time decay—We assumed that the effect of exposure to past affirmative or corrective events (network ties) on the odds of a new correction diminished with the amount of time since the past events occurred. The independent variables capturing network related exposures (described more below) were derived from a characterization of the network of past events—the historical event network—that included a decay on the strength of past events. The historical event network is time-specific and contains all ties occurring before the given time point. The weight of a given past event is equal to $\delta^{(t-1)}$ where t is the number of days in the past the tie last appeared and δ is an estimated value between and potentially including 0 and 1. If $\delta=1$, ties never decay. If $\delta=0$, ties decay to zero immediately. Ties decaying to a weight of less than 0.01 were systematically eliminated from the network prior to deriving the independent variables. We excluded days for which there were fewer than seven days in the past for which affirmation and correction giving was observed. This was a useful rule to account for certain brief periods when TC units temporarily stopped collecting interaction data.

¹This is a subtly different interpretation of a *p*-value than typically allowed with frequentist approaches. A frequentist would have to interpret *p* as the probability of observing the difference we observed assuming that the true difference between the parameters was zero. However, the Bayesian approach we use treats parameters and derived quantities as random variables, the distribution of which can be used to quantify the probability that the parameter sits in a particular range, or the probability that one parameter does not exceed another.

We estimated δ by manually optimizing it. We fit many nearly identical models with the exception that the δ value took on a different value in each model. We reported the model that resulted in the lowest mean deviance value (i.e., -2 times the log likelihood of the data given the parameters) along with the associated value of δ .

Social network predictor variables—The predictors allowed the model to capture how particular configurations of previous events adjust the probability of a new correction occurring within a dyad. These variables can represent complex patterns and can be difficult to name. We therefore adapt a naming scheme from previous social network empirical researchers and use it throughout the remainder of this article (Borgatti, Everett & Johnson, 2013; Ripley, Snijders, Boda, Voros & Preciado, 2016; Wasserman & Faust, 1994). “Indegree” is a term meaning the sum of the weight of affirmations/corrections received by a resident and “outdegree” is the sum for interactions sent by a resident. Independent variables suffixed with “*activity*” adjust the probability of a new correction as a function of a characteristic of the sender (the one creating the *activity*). For example, a variable used in the model to capture the relationship between the number of corrections that the sender has received and the change in odds of sending a correction is “correction-indegree correction-activity.” Independent variables suffixed with “popularity” adjust the probability of the event as a function of a characteristic of the receiver (the one whose *popularity* in receiving corrections is adjusted). For example, “correction-indegree correction-popularity” would capture how recently received corrections adjust the odds of receiving more corrections. Independent variables that capture how similarity or sameness on a characteristic of the sender and receiver adjust the odds of tie between them include “similarity” or “same” in their name. Independent variables that adjust the odds of corrections based on the history of *affirmations* are, in some cases, called *cross-network* variables, or in others include both network types in their name (e.g., “*cross-network reciprocity*” or “*affirmation-indegree correction-activity*”).

Hypotheses and exploratory analyses

Formal hypotheses: Response to staff versus peer affirmation/correction—We hypothesized that the odds of giving a correction will increase more among residents who have recently received peer affirmations or corrections compared with residents who have recently received staff affirmations or corrections. The model includes four variables that capture the adjustment in the odds of giving a correction after receiving an (1) affirmation or (2) correction from a peer, or after receiving an (3) affirmation or (4) correction from staff. These are called (1) *affirmation-indegree correction-activity*, (2) *correction-indegree correction-activity* (3), *staff affirmation-indegree correction-activity* and (4) *staff correction-indegree correction-activity*. To test the hypotheses, we compared the reaction of residents to peer versus staff intervention by comparing the appropriate model parameters with one another (i.e., from those listed above, 1 vs. 3 and 2 vs. 4). The posterior distribution of the model parameters is used to calculate the probability that the effect of a received affirmation from staff is at least as large as the effect of one from a peer. A similar probability is calculated for received corrections, resulting in two hypothesis tests.

Exploratory analysis: Reciprocity—If resident A corrects resident B, it is expected that resident B will be more likely to correct resident A than another resident who did not correct B. There is both theoretical and experimental evidence that reciprocity can stabilize cooperation in large groups (Nowak, 2006; Rand & Nowak, 2013). Reciprocity is common in human social interaction (Falk & Fischbacher, 2006), and it is expected that TC residents will reciprocate help (Stevens, 2013). The model therefore includes a *reciprocity* variable and a cross-network form; *cross-network reciprocity* captures reciprocation of corrections in response to affirmations.

Exploratory analysis: Engagement-based reputation—If peers reward those who are known to cooperate, cooperation in the group as a whole can be more stable (Nowak & Sigmund, 2005; Rand & Nowak, 2013), and there is evidence that reputation effects play a role in TC treatment (Hodge, Warren & Linley, 2014; Miller, Sees & Brown, 2006; Stevens, 2013). It is not clear, however, whether peers will view a reputation for correcting others in a positive or negative light. Reputation variables are called *X-outdegree correction-popularity* (e.g., *correction-outdegree correction-popularity*).

Exploratory analysis: Affirmation/correction sending balance—Another possibility is that TC residents might be more likely to send a correction after sending an affirmation, in order to balance their obligation to the program and possibly avoid a reputation as being unwilling to take the risk of offering negative feedback to peers. The variable capturing this activity balance is *affirmation-outdegree correction-activity*.

Exploratory analysis: Affirmation/correction receiving balance—De Leon (2000) suggests that residents attempt to balance corrections sent to peers with affirmations. Previous research on affirmations confirms this supposition (Doogan & Warren, in press). It is possible that the reverse occurs, that residents balance affirmations with corrections. If true, when resident A affirms resident B, others should be more likely to correct resident B to *communally* balance B's experience (and *affirmation-indegree correction-popularity*). It is also possible that a resident would correct a *specific* peer whom he had previously affirmed, a more focused version of the balance hypothesis. This variable is called *cross-network concentration* ("concentration" being a focus on one individual).

Exploratory analysis: Network structural controls—Social network structural controls were included to account for confounding dependencies in the data that may otherwise confound our hypothesis tests (Pedhazur & Schmelkin, 1991). Several also have clinical implications, so quantifying their contribution to the model is of value.

Transitivity is the possibility that if A corrected B and B corrected C, then A would be more likely to correct C as well. Transitivity is a common characteristic of human social networks that leads to the formation of node clusters and is usually representative of an underlying hierarchy (Wasserman & Faust, 1994).

Uncontrolled autocorrelation is known to increase the likelihood of Type I error (Greene, 2012). Autocorrelation is also of clinical interest as the tendency of a behavior to continue. Four types of autocorrelation are controlled: If A recently sent a correction, he may be more

likely to send a new one (*correction-outdegree correction-activity*). If A recently received a correction, he may be more likely to receive a new one (*correction-indegree correction-popularity*). If A recently corrected B, he is more likely to correct B again (*correction-concentration*). And finally, if A has sent no recent corrections, he will continue not to send them (*correction-isolation correction-activity*).

Exploratory analysis: Individual characteristics—Individual characteristics can influence the peers to whom the resident connects. Homophily, the tendency of individuals to associate with similar others, is the most common pattern (Goodreau, Kitts & Morris, 2009; McPherson, Smith-Lovin & Cook, 2001; Rivera, Soderstrom & Uzzi, 2010). It is of particular concern in TCs because of the possible formation of cliques that could counter the clinical action of the entire community (De Leon, 2000). Residents were expected to be more likely to correct peers of the *same race* (Phillip, 2008), of similar age (*age similarity*), similar LSI-R scores (*LSI-R similarity*), and with similar program entrance dates (*time similarity*).

Individual characteristics can also influence the probability of a correction. The analysis controlled for four individual attributes using both *covariate activity* and *covariate popularity* type variables capturing how covariates adjust the probability of the giving and receiving of corrections. These were age, race (minority code was reference), LSI-R, and time in the program. Older (i.e., more years since birth) residents were expected to send more corrections and receive fewer. It was unclear whether minority residents would send or receive more corrections. It was expected that residents who were at increased risk of recidivism (higher LSI-R scores) would send fewer corrections and receive more. Time in the program was recoded into two dummy variables indicating whether a resident was in his first month of the program or past his fourth month. When both are included, the reference category represents months 2, 3, and 4.

Results

Descriptive statistics are found in Table 1. Seventy-two percent of participants are European-American. The mean age of sampled residents is 28.72 years, the mean score on the LSI-R is 28.14, and mean length of stay is 140.13 days. The mean exposure to received peer intervention is higher than that for staff intervention for both affirmations and corrections.

The results of the model are found in Table 2:

Formal hypothesis tests

Response to peer affirmations and corrections—The *correction-indegree correction-activity* parameter was positive but not statistically significant ($B = .058$, 95% CI $[-.010, .121]$) suggesting residents who receive a correction from a peer are no more or less likely to correct a peer. The *staff correction-indegree correction-activity* parameter was negative but also not statistically significant ($B = -.031$, 95% CI $[-.115, .050]$). While both parameter estimates regarding the response to a received correction are statistically indistinguishable from zero, they *are* statistically distinguishable from one another. The response of residents to staff-sent corrections is unlikely to be as positive as the response to

peer sent corrections ($p = .047$), thus providing evidence confirming our formal hypothesis regarding the differential response to peers versus staff.

Affirmation-indegree correction-outdegree, was statistically significant ($B = .302$, 95% CI [.226, .380]); residents were more likely to give a correction after receiving an affirmation from a peer. *Staff affirmation-indegree correction-activity*, whether residents were more likely to correct a peer after receiving an affirmation from a staff member, was not statistically significant ($B = .062$, 95% CI [−.007, .129]). A comparison of the predicted responses suggests that the response to staff affirmations is exceedingly unlikely to be as positive as the response to peer affirmations ($p < .001$), further confirming our hypothesis that residents will respond more positively to peers than to staff.

Exploratory Analyses

Reciprocity—Residents were more likely to correct a peer who had recently corrected them (*Reciprocity*: $B = 1.145$, 95% CI [.841, 1.443]). Residents did not increase or reduce corrections to peers who affirmed them (*Cross-network reciprocity*: $B = -.012$, 95% CI [−.285, .269]).

Engagement-based popularity—Residents were no more likely to receive a correction after sending a correction (*correction-outdegree correction-popularity*: $B = -.014$, 95% CI [−.087, .060]) or after sending an affirmation (*affirmation-outdegree correction-popularity*: $B = -.013$, 95% CI [−.075, .048]).

Affirmation and correction giving balance—Residents showed an increased tendency to send corrections after having sent affirmations, suggesting an attempt at balance in their own sending of affirmations and corrections (*affirmation-outdegree correction-activity*: $B = .165$, 95% CI [.122, .207]).

Affirmation and correction receiving balance—Residents were not more likely to correct a peer whom they had recently affirmed (*cross-network concentration*: $B = .072$, 95% CI [−.137, .306]). Residents were no more or less likely to correct a peer who had received multiple affirmations from the community (*affirmation-indegree correction-popularity*: $B = .078$, 95% CI [−.010, .176]).

Network structural controls—There is no evidence of transitive closure in giving corrections (*transitivity*: $B = -.006$, 95% CI [−.162, .155]). The *correction-outdegree correction-activity* parameter was positive and statistically significant ($B = .667$, 95% CI [.611, .722]); residents who had recently sent a correction were more likely to send another. The *correction-concentration* parameter was positive and statistically significant ($B = .573$, 95% CI [.351, .781]); those who had recently corrected a specific peer were more likely to correct that peer again. *Correction-indegree Correction-popularity* was positive and statistically significant ($B = .402$, 95% CI [.312, .489]); residents who had recently received a correction were more likely to receive another. Previous isolation with respect to giving corrections was associated with a lower likelihood of giving a correction (*correction-isolation correction-activity*: $B = -.746$, 95% CI [−.906, −.596]).

Individual characteristics—Residents were more likely to correct peers who had spent a similar amount of time in the program (*Time Similarity*: $B = 1.264$, 95% CI [.515, 1.996]). No other characteristic predicted statistically significant homophily in corrections.

European-American residents were more likely to receive a correction than minority residents (*European American popularity*: $B = .154$, 95% CI [.015, .288]), but no more or less likely to give one (*European American activity*: $B = .017$, 95% CI [−.104, .137]). Residents' age and LSI-R scores were unrelated to the odds of giving a correction. The relationship between time spent in the facility and the number of corrections exchanged was nonlinear. Residents were less likely to send corrections in their first month in the program (*first month activity*: $B = −.472$, 95% CI [−.618, −.329]) and after the fourth month (*greater than 4th month activity*: $B = −.598$, 95% CI [−.704, −.495]) than they were in the second, third, and fourth months of their programs. They were also less likely to receive corrections in the first month of their program (*first month popularity*: $B = −.259$, 95% CI [−.358, −.160]) and beyond the fourth month (*greater than 4th month popularity*: $B = −.297$, 95% CI [−.405, −.195]).

Discussion

Residents send significantly more corrections following peer affirmations but do not do so following staff affirmations. Further, while residents did not send statistically significantly more corrections following peer correction, the difference between their tendency to increase the number of corrections sent following peer correction and their tendency to decrease the number of corrections following staff correction was statistically significant. This suggests that residents engage in generalized reciprocity when help comes from peers, but are less likely to do so when help comes from staff.

These findings support the TC use of the community of peers as the method of treatment (De Leon, 2000). They suggest that the best way for TC clinical staff to foster an active unit is to encourage residents to intervene with each other. The increased likelihood of residents to correct peers after receiving a peer affirmation supports De Leon's (2000) claim that affirmations serve to energize residents and increase program participation. Finally, this pattern fits the definition of generalized reciprocity; residents who have received help from peers show a tendency to pass it along (Molm, Collett & Schafer, 2007, Stanca, 2009), perhaps because they feel an increase in their own positive feelings (Isen & Levin, 1972). There is evidence that generalized reciprocity increases feelings of self-competence (Alvarez & van Leeuwen, 2015). Generalized reciprocity in response to peer affirmations may therefore promote positive clinical outcomes.

Peer interaction in sending and receiving corrections is complex. TC residents actively reciprocate peer corrections. This is consistent with TC theory, which claims that residents form connections with peers whom they trust and respect to mutually monitor behavior (Bell, 1994; Stevens, 2013). Such direct reciprocation of help is known to foster cooperation in large groups (Nowak, 2006; Rand & Nowak, 2013) as well as prosocial behavior in individuals (Velez, 2015).

But it is also possible that the exchange of corrections is a form of counterpunishment—residents correct peers who correct them in order to gain revenge (Fehl et al, 2012; Nikiforakis & Englemann, 2011). This analysis cannot separate these explanations, but there are reasons to think that reciprocity more often represents a trusting relationship that allows residents to safely mutually monitor each other. TC clinicians are not fools; they know that residents will sometimes correct peers out of personal animus and take steps to guard against it (Perfas, 2012). Moreover, residents demonstrate homophily by cohort; they were more likely to correct peers who had been in the TC for a similar length of time. This is in line with qualitative evidence that residents who arrive at about the same time develop trust through shared experiences in the TC (Stevens, 2013). This would make it easier and safer to exchange corrections and suggests that the reciprocal exchange is likely a measure of trust rather than revenge.

On the other hand, cohort homophily means that more senior residents may be less likely to correct junior residents. Residents also send and receive more corrections in the middle months of treatment than they do at either the beginning or the end of treatment. TC staff may need to do more to engage senior TC residents in the peer monitoring system, particularly in monitoring junior residents.

Apart from direct reciprocity, residents were no more likely to receive a correction from peers after sending either a correction or an affirmation. We therefore found no evidence that sending either affirmations or corrections creates a reputation to which the community responds. However, autocorrelation in receiving corrections—residents who receive peer corrections on one day are more likely to receive them on the next—may indicate a reputation effect. This may also indicate that residents who break rules on one day are more likely to break them on the next. The analysis controlled for LSI-R score and age, two variables that ought to correlate with residents' propensity to violate TC rules, but these controls are not perfect. To the extent that autocorrelation in the corrections that residents receive does represent a reputation effect it would be possible for residents to find themselves in a vicious circle in which correction leads to further correction. Clinical staff should be aware of this and further research could clarify the matter.

The autocorrelation evident in the number of corrections that residents send means that resident activity in peer monitoring is likely to have a streaky quality—they will engage in a lot for a few days and then relatively little. TC clinicians will therefore have to judge just how active residents need to be in correcting peers, and how long of a break in this form of participation represents a potential problem. The exception is isolates, those who send no corrections in a given time period. They are less likely to send a correction in the future and may need additional encouragement to be willing to mutually monitor peers.

De Leon (2000) commented that affirmations serve to balance corrections. This raises the question of whether residents might also use corrections to balance affirmations. While we did not find evidence that residents correct peers whom they have recently affirmed, they were more likely to send a correction after having sent an affirmation. This raises the possibility that they may be trying to balance their obligations to peers (and possibly their reputations as responsible housemates) by sending both affirmations and corrections.

The analysis failed to find statistically significant homophily (or any statistically significant tendency away from homophily) along lines of age, race, or LSI-R score. This is an encouraging finding, suggesting that residents engage in the peer monitoring system without regard to the demographic characteristics of those whom they are monitoring. In combination with the finding of cohort homophily it suggests that experiences while in the TC are more important in determining the individuals with whom residents exchange corrections than the characteristics which they bring with them.

Among individual characteristics, European Americans were more likely to receive corrections than minority members. Since the analysis controls for homophily by race, this is not the result of homophily in sending corrections by larger numbers of European Americans. Given racial biases in American law enforcement, it is possible that European American correctional TC residents have worse difficulties on average than those of other ethnic groups, and it is possible that the LSI-R does not completely capture these.

Limitations

The data in this analysis has both the strengths and the limitations of archival datasets. It would have been difficult or impossible for researchers to gather a dataset this large, and the fact that the data was recorded as part of ordinary clinical work at the TCs means that researcher bias effects cannot explain the results. On the other hand, the choice of measures was determined by the clinical interests of the facilities rather than the researchers. It is also likely that data entry mistakes led to measurement error (Webb, Campbell, Schwartz & Sechrest, 1972). Finally, the exchange of corrections is only one form of participation. In particular, we do not know the relationship of the affirmations and corrections in this dataset to other behaviors, which limits the clinical conclusions which can be drawn.

Conclusion

Despite these limitations, social network analysis quantifies aspects of the clinical functioning of TCs that would otherwise remain beyond analysis. While some of the current findings, such as the different rates at which residents send corrections following the reception of peer versus staff affirmations, confirm TC clinical theory, others, such as the finding of cohort homophily, have little precedent in the literature. A social network approach to TCs allows researchers use a quantitative framework to grapple with the complexity of the community as method.

Acknowledgments

The authors would like to acknowledge support from NIH grant R21DA023474, and valuable conversations with George De Leon, Ian Hamilton and Jessica Linley.

References

- Alvarez K, Leeuwen E. Paying it forward: How helping others can reduce the psychological threat of receiving help. *Journal of Applied Social Psychology*. 2015; 45(1):1–9. DOI: 10.1111/jasp.12270
- Andrews DA. The level of service inventory-revised. Bonta J, editor Toronto: Multi-Health Systems, Inc; 1995.

- Andrews DA, Bonta J. The psychology of criminal conduct. 5th. Bonta J, editor New Providence, NJ: Matthew Bender & Co; 2010.
- Barr A. Social dilemmas and shame-based sanctions: Experimental results from rural Zimbabwe. Oxford: 2001. Retrieved from
- Bell DC. Connection in therapeutic communities. *International Journal of the Addictions*. 1994; 29(4): 525–543. [PubMed: 8188445]
- Borgatti SP, Everett MG, Johnson JC. Analyzing social networks. Thousand Oaks, CA: Sage; 2013.
- Bowles S, Gintis H. A cooperative species: human reciprocity and its evolution. Princeton: Princeton University Press; 2011.
- Carr WA, Ball SA. Predictors and treatment outcomes of perceived ward atmosphere among therapeutic community residents. *Journal of Substance Abuse Treatment*. 2014; 46(5):567–573. DOI: 10.1016/j.jsat.2014.01.003 [PubMed: 24560437]
- De Leon G. The therapeutic community: Theory, model, and method. New York: Springer; 2000.
- De Nooy W. Networks of action and events over time. A multilevel discrete-time event history model for longitudinal network data. *Social Networks*. 2011; 33(1):31–40.
- Doogan N, Warren K. A network of helping: Generalized reciprocity and cooperative behavior in response to peer and staff affirmations and corrections among therapeutic community residents. In press at *Addiction Research & Theory*.
- Falk A, Fischbacher U. A theory of reciprocity. *Games and Economic Behavior*. 2006; 54(2):293–315. DOI: 10.1016/j.geb.2005.03.001
- Fehl K, Sommerfeld RD, Semmann D, Krambeck HJ, Milinski M, Szolnoki A. I Dare You to Punish Me—Vendettas in Games of Cooperation. *PLoS ONE*. 2012; 7(9):1–7. DOI: 10.1371/journal.pone.0045093
- Fehr E, Gächter S. Altruistic punishment in humans. *Nature*. 2002; 415(6868):137–140. [PubMed: 11805825]
- Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. Boca Raton, Fla: Chapman & Hall/CRC; 2004.
- Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*. 2006; 1(3):515–534. DOI: 10.1214/06-BA117A
- Goodreau SM, Kitts JA, Morris M. Birds of a Feather, or Friend of a Friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks. *Demography*. 2009; 46(1): 103–125. DOI: 10.1353/dem.0.0045 [PubMed: 19348111]
- Greene WH. Econometric analysis. 7th. Boston: Prentice Hall; 2012.
- Hawkins JD, Wacker N. Side bets and secondary adjustments in therapeutic communities. In: De Leon G, Ziegenfuss JT, editors *Therapeutic communities for addictions: Readings in theory, research and practice*. Springfield, Illinois: Charles C. Thomas; 1986. 141–155.
- Herne K, Lappalainen O, Kestilä-Kekkonen E. Experimental comparison of direct, general, and indirect reciprocity. *The Journal of Socio-Economics*. 2013; 45:38–46. DOI: 10.1016/j.soc. 2013.04.003
- Hodge A, Warren K, Linley J. Predictors of resident role model status in therapeutic communities. *Therapeutic Communities: The International Journal of Therapeutic Communities*. 2014; 35(4): 159–167.
- Isen AM, Levin PF. Effect of feeling good on helping: Cookies and kindness. *Journal of Personality and Social Psychology*. 1972; 21(3)
- King G, Zeng L. Explaining Rare Events in International Relations. *International Organization*. 2001; 55(03):693–715.
- Kleidman R. Community organizing and regionalism. *City & Community*. 2004; 3(4):403–421.
- Lerner J, Bussmann M, Snijders TAB. Modeling frequency and type of interaction in event networks. *Corvinus Journal of Sociology & Social Policy*. 2013; 4(1):3–32.
- Makela K, , et al., editors *Alcoholics Anonymous as a mutual-help movement: A study in eight societies*. Madison, WI: University of Wisconsin; 1996. in Press

- Mandell W, Edelen MO, Wenzel SL, Dahl J, Ebener P. Do dimensions of therapeutic community treatment predict retention and outcomes? *Journal of Substance Abuse Treatment*. 2008; 35(3): 223–231. [PubMed: 18248942]
- Masclet D, Noussair C, Tucker S, Villeval MC. Monetary and non-monetary punishment in the voluntary contributions mechanism. *American Economic Review*. 2003; 93(1):366–380.
- McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. *Annual Review of Sociology*. 2001; 27:415–444. DOI: 10.1146/annurev.soc.27.1.415
- Miller S, Sees C, Brown J. Key Aspects of Psychological Change in Residents of a Prison Therapeutic Community: A Focus Group Approach. *Howard Journal of Criminal Justice*. 2006; 45(2):116–128. DOI: 10.1111/j.1468-2311.2006.00409.x
- Molm LD, Collett JL, Schaefer DR. Building solidarity through generalized exchange: A theory of reciprocity. *American Journal of Sociology*. 2007; 113(1):205–242. DOI: 10.1086/517900
- Nikiforakis N, Engelmann D. Altruistic punishment and the threat of feuds. *Journal of Economic Behavior & Organization*. 2011; 78(3)
- Nowak MA. Five rules for the evolution of cooperation. *Science*. 2006; 314:1560–1563. [PubMed: 17158317]
- Nowak MA, Sigmund K. Evolution of indirect reciprocity. *Nature*. 2005; 437:1291–1298. [PubMed: 16251955]
- Oxford House Manual. 7th2015. Downloaded from: <http://www.oxfordhouse.org/userfiles/file/doc/BasMan.pdf>
- Patenaude A. A qualitative exploration into a prison substance abuse treatment program: “I tell them what they want to hear”. In: Sims B, editor *Substance abuse treatment with correctional clients: Practical implications for institutional and community settings*. New York: Haworth Press; 2005.
- Pearce S, Pickard H. How therapeutic communities work: Specific factors related to positive outcome. *International Journal of Social Psychiatry*. 2012; 59(7):636–645. DOI: 10.1177/0020764012450992 [PubMed: 22820178]
- Pedhazur EJ, Schmelkin LP. *Measurement, design, and analysis: an integrated approach*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1991.
- Perfas F. *Deconstructing the therapeutic community: A practice guide for addiction professionals*. Red Hook, NY: Hexagram Publishing; 2012.
- Phillip G. “It’s just black, white or hispanic”: An observational study of racializing moves in California’s segregated prison reception centers. *Law & Society Review*. 2008; 42(4):735–770.
- Rand DG, Nowak MA. Human cooperation. *Trends in Cognitive Sciences*. 2013; 17(8):413–425. [PubMed: 23856025]
- Ripley RM, Snijders TAB, Boda Z, Voros A, Preciado P. *Manual for RSiena*. 2016. Downloaded on 5/27/16 from: https://www.stats.ox.ac.uk/~snijders/siena/RSiena_Manual.pdf
- Rivera MT, Soderstrom SB, Uzzi B. Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms. *Annual Review of Sociology*. 2010; 36:91–115.
- Ronel N, Elisha E, Timor U, Chen G. What do our clients say? Residents’ perceptions of recovery in Retorno—A Jewish therapeutic community. *Addiction Research & Theory*. 2013; 21(4):295–305.
- Stan Development Team. Stan. 2014. Available at: <http://mc-stan.org/>
- Stanca L. Measuring indirect reciprocity: Whose back do we scratch? *Journal of Economic Psychology*. 2009; 30(2):190–202. DOI: 10.1016/j.joe.2008.07.010
- Stevens A. *Offender rehabilitation and therapeutic communities: enabling change the TC way*. London: Routledge; 2013.
- Tiburcio NJ, Kressel D. An Enhanced Understanding of Therapeutic Communities Worldwide. *Revista Addiccion y Ciencia*. 2011; 1(4):1–18.
- Velez JA. Extending the theory of Bounded Generalized Reciprocity: An explanation of the social benefits of cooperative video game play. *Computers in Human Behavior*. 2015; 48:481–491. DOI: 10.1016/j.chb.2015.02.015
- Wasserman S, Faust K. *Social network analysis: methods and applications*. Cambridge; New York: Cambridge University Press; 1994.

Webb EJ, Campbell DT, Schwartz RD, Sechrest L. Unobtrusive measures: Nonreactive research in the social sciences. Chicago: Rand McNally & Company; 1966.

Table 1

Description of residents in the data, and hypotheses oriented time lagged network independent variables in the generalized linear mixed models.

	n	Mean or Percentage	sd	median	min	max
Resident Characteristics (both models)						
<i>Race – European American</i>	1305	0.72	-	-	-	-
<i>Age</i>	1305	28.78	9.25	26	18	61
<i>Level of service inventory - Revised</i>	1305	28.14	7.18	28	5	47
<i>Length of Stay</i>	1305	140.13	36.5	149	0	184
Correction Model Hypothesis Variables						
<i>Res. affirmation-indegree correction-activity</i>	32910	1.41	0.80	1.37	0.00	4.77
<i>Res. correction-indegree correction-activity</i>	32910	0.93	0.71	0.91	0.00	4.09
<i>Staff affirmation-indegree correction-activity</i>	32910	0.32	0.53	0.05	0.00	6.34
<i>Staff correction-indegree correction-activity</i>	32910	0.20	0.44	0.00	0.00	5.98

Table 2

Results of the generalized linear mixed model of corrections.

Predictor Variables	Coefficient	95% CI	
		Lower Bound	Upper Bound
Unit Level Fixed Effects			
rural tc 1	-2.436 *	-2.685	-2.181
rural tc 2	-2.451 *	-2.740	-2.187
urban tc unit 1	-0.969 *	-1.209	-0.732
urban tc unit 2	-1.521 *	-1.778	-1.271
Comparison of Staff to Peer Intervention			
correction-indegree correction-activity	0.058	-0.010	0.121
affirmation-indegree correction-activity	0.302 *	0.226	0.380
staff affirmation-indegree correction-activity	0.062	-0.007	0.129
staff correction-indegree correction-activity	-0.031	-0.115	0.050
Peer Interaction Parameters			
reciprocity	1.145 *	0.841	1.443
transitivity	-0.006	-0.162	0.155
correction-indegree correction-popularity	0.402 *	0.312	0.489
correction-outdegree correction-activity	0.667 *	0.611	0.722
correction-outdegree correction-popularity	-0.014	-0.087	0.060
correction-concentration	0.573 *	0.351	0.781
correction isolation correction activity	-0.746 *	-0.906	-0.596
cross-network reciprocity	-0.012	-0.285	0.269
affirmation-outdegree correction-popularity	-0.013	-0.075	0.048
cross-network concentration	0.072	-0.137	0.306
affirmation-indegree correction-popularity	0.078	-0.010	0.176
affirmation-outdegree correction-activity	0.165 *	0.122	0.207
time similarity (cohort homophily)	1.264 *	0.515	1.996
first month activity	-0.472 *	-0.618	-0.329
first month popularity	-0.259 *	-0.358	-0.160
greater than 4 months activity	-0.598 *	-0.704	-0.495
greater than 4 months popularity	-0.297 *	-0.405	-0.195
age similarity	0.286	-0.034	0.625
same race	0.045	-0.042	0.133
LSI-R similarity	-0.306	-0.655	0.046
age activity ^	-0.006 *	-0.011	0.000
age popularity	-0.001	-0.008	0.007
European American activity	0.017	-0.104	0.137

Predictor Variables	Coefficient	95% CI	
		Lower Bound	Upper Bound
European American popularity	0.154 *	0.015	0.288
LSI-R activity	−0.002	−0.010	0.005
LSI-R popularity	0.009	−0.001	0.019
Random Effect Variances			
sender	0.429	0.376	0.484
receiver	0.663	0.601	0.726
day	0.832	0.750	0.920
Other			
deviance	22821.888	22684.382	22955.846
minimum-deviance decay factor δ	0.790		

* 95% CI did not cross zero; this notation excluded from random effect variances because the intervals cannot cross zero by the definition of variance

^ The upper bound of the age activity 95% CI was very slightly negative before rounding