

Which Helper Behaviors and Intervention Styles are Related to Better Short-Term Outcomes in Telephone Crisis Intervention? Results from a Silent Monitoring Study of Calls to the U.S. 1-800-SUICIDE Network

BRIAN L. MISHARA, PhD, FRANÇOIS CHAGNON, PhD, MARC DAIGLE, PhD,
BOGDAN BALAN, MD, PhD, SYLVAINÉ RAYMOND, MA, ISABELLE MARCOUX, PhD,
CÉCILE BARDON, MA, JULIE K. CAMPBELL, BS, AND ALAN BERMAN, PhD

A total of 2,611 calls to 14 helplines were monitored to observe helper behaviors and caller characteristics and changes during the calls. The relationship between intervention characteristics and call outcomes are reported for 1,431 crisis calls. Empathy and respect, as well as factor-analytically derived scales of supportive approach and good contact and collaborative problem solving were significantly related to positive outcomes, but not active listening. We recommend recruitment of helpers with these characteristics, development of standardized training in those methods that are empirically shown to be effective, and the need for research relating short-term outcomes to long-term effects.

In this article we present data on the effectiveness of telephone help provided to callers in crisis situations (including suicide related crises) and relate characteristics of the process of intervention to short-term outcomes as observed by the end of calls. In the related

article by Mishara et al. (this issue), we presented a detailed description of the nature of telephone interventions observed in crisis calls to the Hopeline Network. We identified two major approaches to telephone help that most centers profess to adopt: active listening

BRIAN MISHARA is Director, Center for Research and Intervention on Suicide and Euthanasia (CRISE) and Professor, Department of Psychology, University of Quebec at Montreal, FRANÇOIS CHAGNON is Associate-director; MARC DAIGLE is Researcher, BOGDAN BALAN is Researcher, SYLVAINÉ RAYMOND is Research coordinator, and ISABELLE MARCOUX is a Researcher at CRISE; CÉCILE BARDON is a PhD student, Department of Psychology, University of Quebec at Montreal; JULIE CAMPBELL is a PhD student, Department of Psychology, University of Quebec at Montreal and Director, Le Faubourg Suicide Prevention Centre, St. Jerome, Quebec; and ALAN BERMAN is Director, American Association of Suicidology.

*This study was conducted under contract with the American Association of Suicidology in fulfillment of the evaluation requirements of Grant No. 6079SM54-27-01-1 from the Substance Abuse and Mental Health Services Administration, U.S. Department of Health and Human Services. Thanks to Reese Butler, the Kristin Brooks Hope Center staff, Jerry Reed, and the Directors and helpers at the crisis centers who participated in this study.

Address correspondence to Brian Mishara, PhD, Director, Center for Research and Intervention on Suicide and Euthanasia, University of Quebec at Montreal, c.p. 8888, Succ. Center-Ville, Montreal, Quebec, Canada, H3C 3P8; E-mail: mishara.brian@uqam.ca

and collaborative problem solving. We presented observations of helper behaviors according to those two models, as well as data on variables identified by center directors as supposedly being more or less helpful. In this paper we examine the relationship between those behaviors by helpers and the short-term outcomes we observed while monitoring calls. We report on a silent monitoring study where the nature of the help provided and caller reactions were observed in 2,611 calls, using standardized instruments, including an assessment of the short-term effects on the callers from the beginning to the end of the calls.

HELPLINE EVALUATIONS: EFFECTIVENESS STUDIES

There is relatively little empirical evidence of helpline effectiveness. Past research focused on three types of effects of suicide prevention helplines: changes in suicide rates in target populations, satisfaction with services, and follow-up or repeated use of services and referrals.

Changes in Suicide Rates in Target Populations

In order to determine if suicide rates within a population have been affected by a suicide prevention activity, it is essential to isolate the effects of the suicide prevention activity from all other significant influences on suicide rates during that time period. Unfortunately, this is both theoretically and empirically impossible to attain (Mishara & Daigle, 2000). Nevertheless, several ingenious methodologies have been used to try to determine if suicide prevention centers have affected suicide rates in local areas. Bagley (1968) found that in a period when suicide rates in England and Wales were declining, rates in towns that had Samaritan activities for at least 2 years declined when compared to other towns that did not have Samaritan helplines available. Yet Kreitman (1976) as-

serted that any declines in the United Kingdom could best be understood as a result of the decreased toxicity of domestic gas, a frequent means of suicide in those years. He found that only suicides by means of toxic gas were significantly reduced, not suicides by other means. Others (Barracough, Jennings, & Moss, 1977) repeated the Bagley study with a more sophisticated matching of experimental and control towns and found no significant differences in suicide rates. However, this more recent study can be criticized because by the time of this study Samaritan services had developed to such an extent that all UK towns had easy access to their services; even if there were no services in their own town, there was always a Samaritan branch available in a neighboring town. Several similarly designed studies also failed to show significant differences (e.g., Bridge, Potkin, Zung, & Soldo, 1977; Leenaars & Lester, 1995, 2004; Lester, 1974; Weiner, 1969).

Telephone helplines should only be effective in helping those who call. This obvious statement implies that simply comparing suicide rates for the entire population of an area where there is a center with those without a center, ignores the fact that one should only find significant impacts among the subpopulation that calls the centers. For example, women are more likely to call than men (Miller, Coombs, Leeper, & Barton, 1984; Mishara & Daigle, 1992; Murphy, Wetzel, Swallow, & McClure, 1969). Miller et al. found significant decreases in suicide rates in U.S. counties that had suicide prevention or crisis centers compared to U.S. counties without those services, but only for White women age 25 and younger, the gender and age group that most often contacted these centers.

Because of the multitude of possible influences on population suicide rates, and the fact that only a small proportion of the population (and only a small proportion of seriously suicidal individuals) actually contact helplines, it is not likely that preventive effects can be convincingly demonstrated by population studies.

Satisfaction with Services

We have not come across a study in which clients were asked to rate their satisfaction with helpline services that did not show highly positive results. When clients are asked to rate their satisfaction or perceived helpfulness, between 60% and 80% give positive responses (e.g., Apsler & Hoople, 1976; King, 1977; McKenna, Nelson, Chatterton, Koperno & Brown, 1975; Motto, 1971; Rogers & Rogers, 1978; Stein & Cotler, 1973; Stein & Lambert, 1984; Streiner & Adam, 1987; Tekavcic-Grad & Zavasnik, 1987; Wold, 1973). One of the reasons satisfaction studies may always get positive results is the fact that response rates in satisfaction studies are generally poor; that is, those who were satisfied are more likely to respond to follow-up than those who were not. Because there are no empirical data linking satisfaction to improvements in clients, and since results are always positive no matter how questions are asked, whom you ask, or what center is conducting the study, satisfaction studies appear to have little value as indicators of the success of helplines.

Repeated Use of Services and Referrals

The fact that clients often call back to a center has been interpreted as an indication that the services are appreciated or helpful. Depending on the follow-up delay, estimates of the proportion of clients who call back vary from 23% to 37% (Apsler & Hoople, 1976; Murphy et al., 1969; Wold, 1973). Mishara and Daigle (1992) estimated that 25% of clients are frequent callers to two Quebec suicide prevention centers, accounting for 63% of all calls. Murphy et al. (1969) used the percentage of persons who showed up for referrals as an indication of effectiveness and reported a 51% no-show rate. Lester (1970) found that no-show rates fluctuated between 44% and 71% for psychotherapy service referrals. There is little way of knowing if people call back because they are getting help or if they call back because they were not helped the first time they

called. Furthermore, compliance with a referral appointment, as important as this appears, does not equate to receiving effective help.

Studies of the Process of Intervention

Several researchers have assumed that if centers have good practices, then they should have good results. This led to investigations on the process of intervention from two perspectives: technical aspects (that is, accomplishing certain tasks during the call), and clinical aspects (qualities of interventions that are considered to be helpful according to a theoretical model). Research on the process of intervention is reviewed in the complementary paper by Mishara et al. (this issue), in which we also present data from the present sample indicating that one can classify telephone help according to an active listening model inspired by the work of Rogers (Rogers, 1951; Truax & Carkhuff, 1967) and a collaborative problem-solving model.

Mishara and Daigle (1997) conducted a study in which they related process measures to assessment of outcomes. They listened to 617 telephone calls from suicidal callers at two primarily French-speaking suicide prevention centers in Canada, categorizing each statement by the volunteers according to a reliable 20-item Helper's Response List which yielded a total of 66,753 responses. Using cluster analytic techniques, intervention styles could be classified as either directive, which included more investigation and direct questions as well as advice and suggestions, or nondirective Rogerian, which consisted of more nondirective and empathetic responses. They found that, within a context where all calls were somewhat directive, having more nondirective Rogerian characteristics was related to a significantly greater decrease in depression, greater likelihood of making a contract with the caller at the end of the call, and greater likelihood of the caller keeping the contract. However, the nature of the call matters. Repeated callers benefited most from a more directive approach, but new callers benefited

significantly more from more non-directive Rogerian techniques.

The results of this study suggest that there may be advantages to certain types of telephone intervention techniques and that these relative advantages may vary according to the nature of the suicidal caller. Linking studies of the process of intervention to call outcomes may help identify which helper behaviors are more likely to result in positive changes in callers.

METHODS

Participants

Of the 91 centers participating in the Hopeline Network at the start of the study, we identified 22 centers that had an average daily call volume above six calls per day, a minimum threshold to justify the cost and effort of monitoring calls and to complete the study within a reasonable period. Eighteen of these centers, whose intervention methods indicated that they currently identified with one of the two intervention models (nondirective and directive), were invited to participate. Fourteen centers agreed to participate and were then asked to have each of their telephone workers be invited to sign and return a consent form guaranteeing anonymity to centers and helpers. For centers in which all the workers did not agree to participate ($n = 6$), we monitored when only consenting workers were working on the lines during that shift. Overall, 782 crisis intervention workers agreed to participate. Eight centers were monitored from August 19 to December 31, 2003; four of these eight centers said they used a directive model and four a nondirective model. Statistical power analyses indicated a total of 400 calls per group would be the minimum necessary to compare the two models, thus our original goal was to monitor a minimum of 800 calls. We continued monitoring with a new group of five centers between January 1 to May 31, 2004. All calls to participating centers received an announce-

ment that calls may be monitored. Another center, which served as a back up for the entire network and received over 40% of all network calls could not be monitored remotely, so a team of three research assistants and a supervisor were sent to this center to monitor calls on site.

Research Assistants

Nine research assistants (RAs), all students in the social sciences, were recruited to monitor calls and received 12 weeks of practice on simulated calls before monitoring began.

Procedures

Two RAs independently monitored each call, one coding observations on the helper and the other coding observations of the caller's behavior. RAs observed all new calls in real time and connected to calls from participating centers over a secure VPN (virtual private network). Calls could then be monitored unobtrusively after entering an access code over a separate line. Data were entered during the call using a computer program that stored all coded information and comments and the time since the start of the call was noted when each data entry was made. In the case of the one center that was monitored on location, RAs similarly monitored calls from a remote location in the center using portable computers.

INSTRUMENTS

Observation of Telephone Helpers' Behaviors

Intervention Process. Helper behaviors were characterized according to individual ratings as well as factor analytically derived factor scores based on the analysis of 26 variables (see Mishara et al., this issue), as follows:

1. Supportive approach and good contact: moral support, good contact, offers call back, reframing, talks about own experience;
2. Active listening: fact questions on the problem, questions on resources, suggests ways to solve the problem, questions on the precipitating events, proposes no-harm contract, suggests plan for action, offers referrals;
3. Collaborative problem solving: reformulation, reflection of feelings, questions on emotions, empowers towards resources, empowers to develop plan of action;
4. Negative style (things not to do): tells caller what to do, reads information, challenges, makes value judgments.

Behavior Ratings at the Conclusion of the Intervention. Center directors felt that by the end of the call there should be one or more of the following elements: an action plan developed for the caller, referrals, agreement on follow-up, and agreement on a contract to not harm oneself.

After the telephone call was completed, RAs coded the helper on two clinical aspects of the intervention following research by Carkhuff (1968) on psychotherapy: empathetic understanding and respect of caller. These are crucial aspects of successful therapeutic intervention according to the active listening model and these skills were used by Lester (1970) in his evaluation of crisis calls. The ratings are on a 5-point Likert scale with detailed definitions provided for each level. To complement these data from the perspective of the more directive collaborative problem-solving model, we developed a similar 5-level rating scale of directivity, ranging from the rating of 1 when the helper "follows completely the rhythm of the caller" (the most nondirective style) to 5 "when the helper completely leads the course of the call" (the most directive method).

Observation of Caller Characteristics

Observations of the callers' behaviors included descriptive information on the

caller, age, sex, and the nature of the current problem that led to the call. The rater indicated if the caller was partially or severely incoherent, and indicated the suspected cause of the incoherence (delusions, alcohol, drugs, handicap, difficulty with English language, and other). In addition, the rater indicated the mood of the caller during the first 3 minutes of the call and at the end of the call, during approximately the last 3 minutes. Since the rater did not know when the call would end, that rating was made immediately after the call ended, based on the observations during the last minutes of the call. Each variable was rated on a 5-point scale with ratings of 1 and 5 indicating the caller behavior matched the definition of the contrasting states observed, ratings of 2 and 4 indicated the caller tended to resemble the definition of the item, and a rating of 3 indicated that one could not determine which of the opposing emotions best described the state of the caller. Detailed definitions were provided for each scale and raters were trained with role-plays and by listening to fictional and real calls to develop reliability of the ratings. The scales were: apprehensive–confidence; sad–happy; helpless–resourceful; tired–dynamic; hopeless–hopeful; confused–decided. In addition, four emotional states were assessed as being present, somewhat present, or absent at the beginning and at the end of the call: crying, desperate, depressive mood, and agitated. These unidimensional ratings were developed after attempts to pair these items with an opposite failed to yield reliable ratings in pre-tests. We also rated if the caller was ambivalent about suicide by indicating if caller: "is ambivalent about suicide," "wants to live," "wants to die," or "there is not enough information available to say either way."

The rater indicated if the caller was calling to obtain a specific referral or information. At the end of the call, three items indicated if there was agreement between the caller and helper on a no-harm contract, agreement on follow-up, and agreement on the caller calling back. After the call ended, the raters completed the Crisis Call Outcome Rating Scale (CCORS) developed by Bonne-

son and Hartsough (1987). This 26-item scale has 16 items to indicate successful outcomes and 10 that indicate unsuccessful outcomes, rated on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*). The scale score can range from 26 to 182. The content of the scale overlapped partially with some other data we collected but provides more specific information and has been used in other studies on telephone help with suicidal callers.

Inter-rater Reliability

Nine separate RAs independently coded 57 calls during two periods: the first after completing their training, the second after final adjustments were made to the coding categories and their definitions during the first 2 weeks of actual call monitoring. This procedure was repeated, but instead of having all nine raters simultaneously rate the calls, four persons were scheduled to listen independently to each call, two rating the helper behavior and two rating the caller behavior. No discussions of ratings were made and raters conducted their ratings independently. At the beginning of listening to calls the percentage agreement was calculated for 99 calls. A little over half way through the study another 99 calls were monitored with two RAs again simultaneously rating the helpers and two others simultaneously rating the callers.

Overall reliability was quite high for both helper behaviors (see Mishara et al., this issue) and observations of caller behaviors (see Table 1). Note that in Table 1 the percentage of absolute agreement on items that were generally rated on 5-point Likert scales are shown. Thus the absolute percentage constitutes a conservative estimate of reliability; when there were disagreements they were generally between ratings in the same direction.

Ethical Concerns

This research study was approved by the AAS Institutional Review Board (IRB) constituted at the University of Chicago as well as the IRB of the University of Quebec at Montreal. Callers were informed the calls

may be monitored over a recorded announcement that was played before callers were connected to participating centers and all helpers who were monitored signed consent forms (see Mishara et al., this issue, for a more detailed discussion of ethical aspects of this study). No recordings of any calls were made and all ratings were undertaken in real time.

RESULTS

Calls Included and Excluded

For the purpose of this report, we retained all of the calls that we were able to monitor and which concerned a crisis situation (1,431 of the 2,611 calls [54.8%] monitored; see Mishara et al., this issue, for a more detailed presentation of call inclusion and exclusion criteria).

The average length of calls in the sample we retained was 18.8 minutes, with a standard deviation of 16.3 minutes. Call lengths varied from 3 to 162 minutes. Calls from persons in suicidal crises lasted longer than calls from third-party callers and persons in crises that did not concern suicide ($F = 39.04$, $df = 4, 1369$, $p < .001$); the longest calls were with persons in an attempt and persons who had already planned what method to use.

The ratio of calls from women and men is about 3 to 2, except that the third-party callers had a ratio closer to 3 to 1. In this study, we did not monitor any call if there was an indication or suggestion that the person was below 18 years of age. Since the age was not always asked, we were not always able to have an accurate age; however, approximate age indication shows that 47% of callers were between 18 and 34, 44% between 35 and 54, 6% between 55 and 65, and only 1% 65 years and older.

Call Outcomes

Overall, many variables did not significantly change from the beginning to the end of the call according to our observations.

TABLE 1
*Reliability Analyses: Average Percentage of Interrater Agreement (8 raters)
 on Ratings of Caller Behavior*

Item	Average % of Agreement		
	Before study (<i>n</i> = 57 calls)	Beginning of study (<i>n</i> = 99 calls)	Middle of study (<i>n</i> = 99 calls)
Apprehensive/Confident—initial	82	91.7	83.8
Sad/Happy—initial	69	77.8	86.7
Tired/Dynamic—initial	57	58.3	80.8
Helpless/Resourceful—initial	73	61.1	84.8
Hopeless/Hopeful—initial	78	72.2	85.8
Confused/Decided—initial	58	55.6	81.8
Crying—initial	100	96.3	92.9
Depressive mood—initial	80	74.1	81.8
Desperate—initial	91	94.4	82.8
Agitated—initial	96	83.3	88.7
Ambivalent—initial	82	90.7	89.9
Apprehensive/Confident—final	60	82.7	92.1
Sad/Happy—final	77	82.7	88.7
Tired/Dynamic—final	71	82.7	85.4
Helpless/Resourceful—final	75	51.7	82.9
Hopeless/Hopeful—final	75	65.5	80.9
Confused/Decided—final	77	58.6	91.0
Crying—final	64	93.3	95.5
Depressive mood—final	96	84.4	89.7
Desperate—final	77	95.5	80.6
Agitated—final	82	84.4	94.4
Ambivalent—final	94	84.1	94.3
Acceptance of No Harm Contract	80	95.1	93.1
Agreement on follow up	87	95.1	91.1
Agreement on caller call back	91	87.8	89.1

When change did occur, it was much more likely to be in the direction of improvement. Table 2 shows the number and percentage of calls in which there was improvement, no change, or the caller feeling worse at the end of the call. Over half the callers were rated as less confused and more decided by the end of the call, almost half were rated more resourceful and less helpless, 40% were rated more hopeful, and 38% were rated more confident. However, 11% of callers were rated more apprehensive at the end of the call, 9% sadder, 9.5% more helpless, 11% more hopeless, and 10% more confused.

Centers varied significantly in the extent that there were positive and negative

changes between the beginning and end of the calls, according to repeated measures MANOVA in which the independent variable was the 14 participating centers and the dependent variables were all 11 within-subjects ratings that were made at the beginning and at the end of the calls. There was a significant mean effect of changes overall from the beginning to the end of the call ($F = 31.97$; $df = 10, 1197$; $p < .001$), and there was an interaction between the independent variable of differences between the centers and the pre-post effects ($F = 1.352$; $df = 130, 1206$; $p < .05$).

We conducted a MANOVA test in which the dependent variables were the re-

TABLE 2
Outcomes for the Overall Sample

Outcomes	Direction of Change							
	Decrease ^a		No change ^b		Improvement ^c		Missing Data	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Apprehensive/Confident	160	11.2	705	49.2	548	38.3	19	1.3
Sad/Happy	131	9.1	965	67.4	318	22.2	18	1.3
Tired/Dynamic	211	14.7	961	67.1	241	16.8	19	1.3
Helpless/Resourceful	136	9.5	581	40.6	698	48.7	17	1.2
Hopeless/Hopeful	159	11.1	677	47.3	575	40.2	21	1.5
Confused/Decided	142	9.9	520	36.3	749	52.3	21	1.5
Crying	52	3.6	1,207	84.3	154	10.8	19	1.3
Depressive mood	94	6.6	1,058	73.9	262	18.3	18	1.3
Desperate	89	6.3	1,098	76.7	225	15.7	20	1.4
Agitated	66	4.6	1,190	83.1	155	10.8	21	1.5
Ambivalent	24	1.7	1,006	70.3	165	11.5	237	16.6

^aDecrease: there was a change toward the worst along this dimension

^bNo change: there was no change measured between the beginning and end of call along this dimension

^cImprovement: there was a change towards an improvement along this dimension

peated measures of the different ratings of states of the caller at the beginning and end of the call and the independent variable was type of calls classified as suicidal crisis, non-suicidal crisis, and third-party suicide related calls. There was an overall significant within subject change from the beginning to end of calls ($F = 72.33$; $df = 10, 1384$; $p < .001$) and a significant interaction between the type of crisis situation and changes from the beginning to the end of the call ($F = 7244.28$; $df = 10, 1384$; $p < .001$). Univariate tests showed significant positive changes from the beginning to the end of each call on all the variables except tired–dynamic, with callers being rated as being on the average more tired at the end of the call, although the change from beginning to end of calls was not significant. Changes in the following variables were not related to the type of crisis situation in univariate interactions: sad–happy, helpless–resourceful, tired–dynamic, hopeless–hopeful, and agitated–less agitated. There was an interaction between changes from the beginning and end of call and the type of

caller on the variables of apprehensive–confident, confused–decided, crying, depressive mood, and feeling desperate. In the cases of apprehensive–confident and sad–happy, there were more improvements in callers experiencing a suicidal crisis. Helplessness increased more with callers in a nonsuicidal crisis. Suicidal callers tended to be sadder at the end of the call whereas third-party callers appeared to be less tired and improved significantly more than suicidal callers in terms of being less confused by the end of the call. Nonsuicidal crisis callers were crying less than the other two groups by the end of the call and there were improvements in depressive moods in suicidal and nonsuicidal crises, but not in third-party callers (perhaps due to a ceiling effect, since they were rarely rated as being depressed). Nonsuicidal crisis callers improved most in decreasing feelings of being desperate.

The mean Crisis Call Outcome Rating Scale (CCORS) score for the total sample was 102.4 ($SD = 18.51$). Centers had CCORS scores ranging from an average of 97.61 to

109.36. ANOVA comparing centers in their CCORS scores shows significant differences between centers ($F = 1.87$; $df = 13, 1226$; $p < .03$)

Relationships Between Intervention Styles and Outcomes

MANOVA analyses were conducted in which the dependent variables were the callers' behavioral ratings at the beginning and end of the calls and the independent variable was scores on the four factors of helpers' behaviors: supportive approach and good contact, active listening, collaborative problem solving, and negative approach. There was a significant interaction between pre-post changes and supportive approach and good contact ($F = 88.90$; $df = 9, 1366$; $p < .001$). Univariate tests indicated significant positive changes related to supportive approach and good contact on the variables: sad-happy ($p < .025$), helpless-resourceful ($p < .001$), hopeless-hopeful ($p < .001$), confused-decided ($p < .001$), crying (decreased) ($p < .001$), depressive mood (decreased) ($p < .018$), and desperate (decreased) ($p < .001$). Active listening was only significantly related in univariate analyses to a significant change of having more crying ($F = 7.10$, $df = 1, 1376$, $p < .008$).

Collaborative problem solving also had a significant interaction with the pre-post variables in MANOVA ($F = 6918.70$; $df = 9, 1352$; $p < .001$). Univariate tests indicated significant positive changes in helpless-resourceful ($p < .001$), hopeless-hopeful ($p < .001$), confused-decided ($p < .001$), crying (decreased) ($p < .001$), depressive mood ($p < .018$), and desperate (decreased) ($p < .001$). Negative approach was not significantly related to pre-post changes in the MANOVA and no univariate tests showed significant relationships.

The ratings on the CCORS was significantly correlated with the number of improvements in the helpers' behavioral rating ($r = .50$, $p < .001$). When relating the four styles identified in the factor analyses with CCORS scores for the entire sample, it was found that supportive approach and good

contact as well as collaborative problem solving were related to higher CCORS scores ($r^2 = .10$). When only the 503 suicidal callers were considered, the same two helper styles were related to CCORS, but a higher percentage of the variance was explained ($r^2 = .14$). However, with the 215 high-risk suicidal callers only supportive approach and good contact was significantly related to CCORS scores ($r^2 = .13$).

There are several other possible indications of improvement which were analyzed. We conducted stepwise logistic regression analyses of helper styles and the outcome of reaching an agreement for a no-harm contract by the end of the call for the entire sample as well as for sub samples of caller categories. All four helper styles were related to reaching an agreement by the end of the call for the entire sample; however, as one might predict, the negative style was negatively related. When only suicidal callers were included in the analysis, the only variable retained in the regression analysis was supportive approach and good contact (OR = 1.78; 95% CI = 1.481-1.993). Also, only supportive approach and good contact was retained when only high risk suicidal callers were included in the analysis (OR = 1.76; 95% CI = 1.397-2.221).

A likely indication of a negative outcome is if the caller hangs up on the helper before the call has ended. Both supportive approach and good contact and collaborative problem solving are negatively related to caller hang ups in the suicidal sample (OR = .558; 95% CI = .444-.702 for supportive approach and good contact; OR = .794; 95% CI = .640-.985 for collaborative problem solving). When the entire sample was considered, only supportive approach and good contact were related (OR = .501; 95% CI = .358-.702). When only high risk suicidal callers were considered, only supportive approach and good contact was retained in the equation.

Ratings of Empathy, Respect, and Directivity

The observer's rating on empathy was significantly related to the CCORS, as indi-

cated in a one-way ANOVA test in which helpers were divided into three levels of empathy: low (levels 1 and 2), average (level 3), and high (levels 4 and 5) ($F = 60.74$, $df = 2$, 1413; $p < .001$). There is a very strong and significant association between empathy and caller hang ups. Overall, 115 of the callers (8%) hung up on helpers. Helpers with low empathy had a 12.5% hang up rate, helpers with average empathy had a 6.6% hang up rate, and helpers with high empathy ratings had only a 2.4% hang up rate ($\chi^2 = 22.44$, $df = 2$; $p < .001$).

Empathy was also related significantly to reaching an agreement by the end of the call. Low empathy helpers reached an agreement in 41.8% of calls, compared to average empathy calls, which ended with an agreement rate of 58.4%; high empathy calls ended in an agreement in 73.3% of calls.

Repeated measures MANOVA were conducted using empathy levels as the independent variable and the callers' behavioral observations at the beginning and end of the call as the dependent variables. There was a significant interaction between empathy and behavior ratings overall ($F = 5.25$; $df = 18$, 1395 $p < .001$). Higher empathy, according to univariate tests, was significantly related to improvements in apprehensive–confident ($p < .001$), sad–happy ($p < .001$), helpless–resourceful ($p < .01$), hopeless–hopeful ($p < .001$), confused–decided ($p < .001$), crying (decreased) ($p < .001$), depressive mood ($p < .05$), and desperate ($p < .001$).

Ratings on respect, coded as high, average, and low in the same manner as empathy, were also significantly related to the CCORS results ($F = 56.12$; $df = 2$, 1413; $p < .001$), with more respect being related to higher CCORS scores. Ratings on respect were related to caller hang ups; the higher the level of respect the fewer the hang ups ($\chi^2 = 17.94$, $df = 2$; $p < .001$). Helpers with lower respect had rates of 14.2% hang ups; helpers with average respect had a hang up rate of 7.2%, and helpers with high respect had hang ups in only 4.1% of calls. Respect was also significantly related to reaching an agreement by the end of the call ($\chi^2 = 24.57$, $df = 2$; $p < .001$). Calls in which there was low

respect ended in an agreement 33.3% of the time, calls with average respect ended in an agreement in 59.8% of calls, and high respect calls ended in agreement 66% of the time.

Repeated measures MANOVA test results indicated a significant interaction between respect and the dependent variables ($F = 5.43$, $df = 18$, $p < .001$). Significant variables related to respect include improvements in: apprehensive–confident ($p < .001$), sad–happy ($p < .001$), helpless–resourceful ($p < .001$), hopeless–hopeful ($p < .001$), confused–decided ($p < .001$), crying (decreased) ($p < .001$), depressive mood ($p < .001$), and desperate ($p < .001$).

Relationship Between Ratings of Directivity and the Outcome Variables

Ratings on directivity consisted of 5-point ratings from “Helper follows Completely the Rhythm of the Caller” (=nondirective) to “Helper completely leads the course of the call” (=directive). These five ratings were subsequently re-coded as directive (scores 4 and 5), mixed (score of 3), and nondirective (scores of 1 and 2). ANOVA indicated a significant relationship to the CCORS results ($F = 25.910$; $df = 2$, 1412; $p < .001$). As shown in Figure 1 and confirmed in *post hoc* tests, there were significant differences between all levels of directivity, with the mixed calls being related to the highest CCORS scores, followed by the directive ones.

There were no significant relationships between ratings on directivity and caller hang ups, or reaching an agreement by the end of the call.

Ratings on directivity were associated significantly ($F = 2.63$; $df = 9$, 1394, $p < .001$) with changes from the beginning to the end of the call on the behavioral rating, as indicated by the significant MANOVA interaction between directivity and the dependent variables. In univariate tests, significant changes were observed in helpless–resourceful ($p < .001$), confused–decided ($p < .001$), and desperate ($p < .045$). *Post hoc* analyses indicated that the mixed directivity calls are related

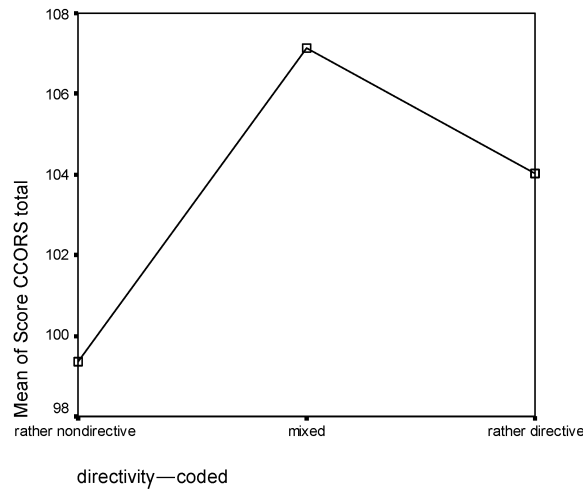


Figure 1. CCORS Scores in Relationship with Directivity

more significantly to positive outcomes, followed by the directive ones.

DISCUSSION AND CONCLUSIONS

The data gathered on the intervention process are derived from theoretical models that are based on past research and interviews with center directors. They were then subjected to an intensive process of refinement and assessment of inter-rater reliability. The belief in the relationship between the short-term effects we observed and long-term benefits requires a leap of faith since little empirical proof exists that positive changes at the end of a call are related to long-term positive outcomes. The only way to ascertain if short-term effects are related to long-term benefits is to conduct a longitudinal investigation with a control group. Such a study would have to follow callers for a sufficiently long period of time in order to determine that any benefits are not just temporary. A control group would have to be used to make certain that the benefits are not the result of spontaneous improvement. Since callers are affected by a multitude of life circumstances and may engage in other help-seeking activities, it is a great challenge to differentiate effects that are related to telephone interven-

tions from the effects of other activities. This type of longitudinal research has never been conducted and, because of the confounding factors involved in real life studies, may be technically impossible to undertake. In the meantime, we must realize the possible limitations of short-term outcome measures and accept the fact that our interpretations are based on a trust in their theoretical roots. We must also maintain a healthy level of conservatism in generalizations from our findings.

We found that positive call outcomes are more likely when the helper expresses empathy and respect for the callers. The ability to be empathetic and communicate respect may be qualities of helpers rather than skills that may be easily taught as part of a training program. Chad Varah said this when he founded the Samaritan movement in England over 40 years ago (personal communication, BLM). He felt that some people are natural "befrienders" and others are not. The clear relationships we observed between these qualities and positive effects suggest that these are qualities that centers should screen for in their telephone helpers.

The most powerful predictors of positive outcomes were the items in the factor analytically derived scale of supportive approach and good contact. These behaviors include: validation of emotions, giving moral

support, good contact, reframing, talking about own experience, and offers to call back. Training activities exist to teach these behaviors and our results indicate that they should be encouraged.

Collaborative problem solving, one of the four factor analytically derived variables describing the process of intervention, is significantly related to positive outcomes. Collaborative problem solving involves asking fact questions on the problem, questioning about resources, suggesting ways to solve the problem, questions on precipitating events, proposing a no-harm contract, suggesting a plan of action, and offering referrals. It is important to note that all helper behaviors we studied were based on the center directors' descriptions of practices in their centers. These helper behaviors are not necessarily evidence-based practices, and the use of some, such as proposing a no-harm contract, may be controversial.

It is important to note that supportive attitude and good contact explains more of the variance in outcomes than collaborative problem solving. Nevertheless, in the more qualitative analyses that involved observer ratings of directivity (the extent to which helpers follow the caller or lead the conversation), it is the mixed style that combines directive and nondirective techniques that was related to more positive outcomes.

There were a number of things that center directors all agreed helpers should not do. These included telling the caller what to do, reading information, challenging the caller, making value judgments, and moralizing. These behaviors all occurred sometimes, although infrequently. There are a few examples when these behaviors had a negative effect, but overall, these behaviors were not significantly related to positive or negative outcomes according to our results, with one notable exception: Helpers talking about their own experiences was related to positive outcomes. Sharing personal experiences appears to be helpful to callers in crisis, providing that they are used to convey an understanding of the caller's situation. The prescription of not sharing personal experiences

is derived from theories of psychotherapy and may not be relevant to telephone crisis intervention, although this warrants further investigation. Telephone interventions may not need to maintain as strict a boundary between the helper and the caller as face-to-face psychotherapy, where the client sees the therapist and the helper is identified. Helpers sharing their experiences on the phone may help compensate for the anonymity of the contact on a helpline.

Recommendations

Our findings highlight the need for better quality control of the nature of telephone interventions. There is currently no monitoring or control on the Hopeline Network to determine if their practices meet minimal standards of good practice or if callers appear to be helped by the end of the call. One of the most simple and effective means of guaranteeing that callers receive good and appropriate help would be to require that calls to centers are monitored for quality assurance purposes. This is already current practice in most commercial services offered over the telephone.

Our findings indicate that centers should look for the qualities of empathy and respect, and the ability to establish a good initial contact, in telephone helpers and develop selection procedures and criteria that take these qualities into consideration.

There is no standard training curriculum on crisis intervention with suicidal callers. Each center develops its own training content and methods, though often based upon pre-existing models or training programs adapted to local needs. Our findings suggest specific characteristics of telephone helper behaviors that are related to more positive call outcomes, particularly the need to establish a good contact and the use of the collaborative problem solving approach. It may be useful for a central organization, such as the AAS, to develop model training programs that take into consideration these findings and are subjected to careful evaluations of their effectiveness.

One of the greatest challenges in future research on telephone help is to develop means for determining the relationship between short-term proximal effects which one may observe during unobtrusive observations of calls and the long-term impact on callers. Although the methodological challenges may

appear daunting, it is important to conduct long-term longitudinal investigations in order to better understand if telephone suicide prevention helplines actually prevent suicidal behavior and have significant positive effects on people's lives.

REFERENCES

- APSLER, R., & HOOPLE, H. (1976). Evaluation of crisis intervention services with anonymous clients. *American Journal of Community Psychology*, 4, 293-302.
- BAGLEY, C. R. (1968). The evaluation of a suicide prevention scheme by an ecological method. *Social Science and Medicine*, 2, 1-14.
- BARRACLOUGH, B. M., JENNINGS, C., & MOSS, J. R. (1977, July 30). Suicide prevention by the Samaritans: A controlled study of effectiveness. *The Lancet*, 237-239.
- BONNESON, M. E., & HARTSOUGH, D. M. (1987). Development of the Crisis Call Outcome Rating Scale. *Journal of Consulting and Clinical Psychology*, 55, 612-614.
- BRIDGE, T. P., POTKIN, S. G., ZUNG, W.W.K., & SOLDI, B. J. (1977). Suicide prevention centers: Ecological study of effectiveness. *The Journal of Nervous and Mental Disease*, 164, 18-24.
- CARKHUFF, R. R. (1968). Differential functioning of lay and professional helpers. *Journal of Counseling Psychology*, 15, 117-126.
- KING, G. D. (1977). An evaluation of the effectiveness of a telephone counseling center. *American Journal of Community Psychology*, 5, 75-83.
- KREITMAN, N. (1976). The coal gas story: United Kingdom suicide rates, 1960-71. *British Journal of Preventive and Social Medicine*, 30, 86-93.
- LEENAARS, A. A., & LESTER, D. (1995). Impact of suicide prevention centers on suicide in Canada. *Crisis*, 16, 39.
- LEENAARS, A., & LESTER, D. (2004). The impact of suicide prevention centers on the suicide rate in the Canadian provinces. *Crisis*, 25, 65-68.
- LESTER, D. (1970). Steps toward the evaluation of a suicide prevention center (parts 1 to 4). *Crisis Intervention*, 2, 42-45.
- LESTER, D. (1974). Effect of suicide prevention centers on suicide rates in the United States. *Health Services Reports*, 89, 37-39.
- McKENNA, J., NELSON, G., CHATTERSON, J., KOPERNO, M., & BROWN, J. H. (1975). Chronically and acutely suicidal persons one month after contact with a crisis intervention center. *Canadian Psychiatric Association Journal*, 20, 451-454.
- MILLER, H. L., COOMBS, D. W., LEEPER, J. D., & BARTON, S. N. (1984). An analysis of the effects of suicide prevention facilities on suicide rates in the United States. *American Journal of Public Health*, 74, 340-343.
- MISHARA, B. L., CHAGNON, F., DAIGLE, M., BALAN, B., RAYMOND, S., MARCOUX, I., ET AL. (this issue). Comparing models of helper behavior to actual practice in telephone crisis intervention: A silent monitoring study of calls to the U.S. 1-800-SUICIDE network. *Suicide and Life-Threatening Behavior*, 37, 000-000.
- MISHARA, B. L., & DAIGLE, M. (1992). The effectiveness of suicide prevention centers. *Santé Mentale au Canada/Canada's Mental Health*, 40(3), 24-29.
- MISHARA, B. L., & DAIGLE, M. (1997). Effects of different telephone intervention styles with suicidal callers at two suicide prevention centers: An empirical investigation. *American Journal of Community Psychology*, 25, 861-895.
- MISHARA, B. L., & DAIGLE, M. (2000). Helplines and crisis intervention services: Challenges for the future. In D. Lester (Ed.), *Suicide prevention: Resources for the millennium* (pp. 153-171). Philadelphia: Brunner/Mazel.
- MOTTO, J. A. (1971). Evaluation of a suicide prevention center by sampling the population at risk. *Suicide and Life-Threatening Behavior*, 1, 18-22.
- MURPHY, G. E., WETZEL, R. D., SWALLOW, C. S., & McCLURE, J. N. (1969). Who calls the suicide prevention center: A study of 55 persons calling on their own behalf. *American Journal of Psychiatry*, 126, 314-324.
- ROGERS, C. R. (1951). *Client-centered therapy*. Boston: Houghton Mifflin.
- ROGERS, R. L., & ROGERS, C. A. (1978). An analysis of telephone crisis intervention counseling based upon consumer evaluation. *Crisis Intervention*, 9, 102-116.
- STEIN, C. M., & COTLER, S. (1973). Crisis phone services: Evaluation of a hotline program. *American Journal of Community Psychology*, 12, 101-126.

- STEIN, D. M., & LAMBERT, M. J. (1984). Telephone counseling and crisis intervention: A review. *American Journal of Community Psychology, 12*, 101–126.
- STREINER, D. L., & ADAM, K. S. (1987). Evaluation of the effectiveness of suicide prevention programs: A methodological perspective. *Suicide and Life-Threatening Behavior, 17*, 93–106.
- TEKAVCIC-GRAD, O., & ZAVASNIK, A. (1987). Comparison between counselor's and caller's expectations and their realization on the telephone crisis line. *Crisis, 8*, 162–177.
- TRUAX, C. B., & CARKHUFF, R. R. (1967). *Toward effective counseling and psychotherapy: Training and practice*. Chicago: Aldine.
- WEINER, I. W. (1969). The effectiveness of a suicide prevention program. *Mental Hygiene, 53*, 357–363.
- WOLD, C. I. (1973). A two-year follow-up of suicide prevention center patients. *Suicide and Life-Threatening Behavior, 3*, 171–183.

Manuscript Received: July 28, 2006

Manuscript Accepted: November 1, 2006