PSYCHOLOGICAL AUTOPSY STUDIES AS DIAGNOSTIC TOOLS: ARE THEY METHODOLOGICALLY FLAWED?

HEIDI HJELMELAND
Department of Social Work and Health Science, Norwegian University of Science and Technology, Trondheim and Department of Health Surveillance and Suicide Prevention, Norwegian Institute of Public Health, Oslo, Norway

GUDRUN DIESERUD
Department of Health Surveillance and Suicide Prevention, Norwegian Institute of Public Health, Oslo, Norway

KARI DYREGROV
Department of Health Surveillance and Suicide Prevention, Norwegian Institute of Public Health, Oslo and Center for Crisis Psychology, Bergen, Norway

BIRTHE L. KNIZEK
Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway

ANTOON A. LEENAARS
Department of Health Surveillance and Suicide Prevention, Norwegian Institute of Public Health, Oslo, Norway and Windsor, Canada

One of the most established “truths” in suicidology is that almost all (90% or more) of those who kill themselves suffer from one or more mental disorders, and a causal link between the two is implied. Psychological autopsy (PA) studies constitute one main evidence base for this conclusion. However, there has been little reflection on the reliability and validity of this method. For example, psychiatric diagnoses are assigned to people who have died by suicide by interviewing a few of the relatives and/or friends, often many years after the suicide. In this article, we scrutinize PA studies with particular focus on the diagnostic process and demonstrate that they cannot constitute a valid evidence base for a strong relationship between mental disorders and suicide. We show that most questions asked to assign a diagnosis are impossible to answer reliably.
by proxies, and thus, one cannot validly make conclusions. Thus, as a diagnostic tool psychological autopsies should now be abandoned. Instead, we recommend qualitative approaches focusing on the understanding of suicide beyond mental disorders, where narratives from a relatively high number of informants around each suicide are systematically analyzed in terms of the informants’ relationships with the deceased.

Psychological autopsy (PA) was originally a method to clarify the mode of death in cases where it was equivocal by examining, in detail, factors surrounding the death and the deceased (Shneidman, 1981). Somewhere along the way, however, this method has become the prime approach in studying risk factors for suicide. Presently, PA is regarded as the most direct (Cavanagh, Carson, Sharpe, & Lawrie, 2003), as well as a reliable and valid, method (Kelly & Mann, 1996) to study the relationship between various explanatory factors and suicide. Often, mental disorders, particularly mood disorders, are presented as the most relevant antecedent of suicide and often a causal link between mental illness and suicide is implied (e.g., Cavanagh et al., 2003; Isacsson & Rich, 2003).

In Cavanagh and colleagues’ (2003) comprehensive review of PA studies, they concluded that around 90% of those who die by suicide suffer from one or more mental disorders. In terms of depression being found in the majority of suicides, Isacsson and Rich (2003) claimed that, “This finding has been replicated over and over again and we believe that many, like us, have concluded that this connection has been replicated enough to be proven” (p. 457). In the words of Berman (2006), “Sometimes research findings, particularly if replicated, quickly achieve the status of fact. One of these, established through PA studies, is that on average 90 percent or more of those who died by suicide had a retrospectively diagnosable mental disorder” (p. 3). Based on replication, this now seems to be an established “truth” among suicidologists, in spite of the numerous and serious methodological problems inherent in PA studies that have been demonstrated (e.g., Beskow, Runeson, & Åsgård, 1990; Hawton et al., 1998; Pouliot & De Leo, 2006). This “truth” has had immense consequences because it has contributed enormously to the current high emphasis on identification and treatment of mental disorders as the main suicide preventive effort (Bertolote, Fleischman, De Leo, & Wasserman, 2003; Cavanagh et al., 2003; Mortensen, Agerbo, Erikson, Qin, & Westergaard-Nielsen, 2000; Portzky, Audenhart, & Van Heeringen, 2005).
However, with the large number of PA studies (a search on psychological autopsy in MEDLINE in January 2010 yielded 457 hits) with all their serious methodological weaknesses, a relevant question is whether we here have a case of mere number constituting the “evidence base.” In our opinion, the severe methodological problems inherent in PA studies have not received the attention they deserve. Because PA studies are the most important contributor to the “evidence base” for the strong relationship between mental disorders and suicide, we need to scrutinize how the information is gathered and how psychiatric diagnoses are assigned in such studies. Previous studies investigating the reliability and validity of PA studies may have missed one of the most fundamental issues in research, namely the importance of choosing a method that is appropriate to answer the research questions. The relevant and very important question here is: Is it really possible to assign psychiatric diagnoses to someone who is dead by interviewing someone else?

In this article, we will demonstrate why PAs in the most common form today cannot constitute a valid evidence base for a strong relationship between psychiatric disorders (per se) and suicide. We do this by scrutinizing, in more detail than what has been done before, how psychiatric diagnoses are assigned to the deceased in such studies. But first, a brief recapitulation of the methodological problems pointed to by others.

**Methodological Problems in PAs Revisited**

Over the years, a number of serious methodological problems inherent in PA studies have been discussed (e.g., Beskow et al., 1990; Hawton et al., 1998; Pouliot & De Leo, 2006). In a particularly meticulous paper, Pouliot and De Leo (2006) listed the following serious weaknesses: (a) Most PAs are conducted under the medical model paradigm and a causal link between a mental disorder and suicide is drawn. However, this model does not fully account for the fact that psychopathology never is a sufficient cause of suicide, although it might be a contributory one. Moreover, suicide rates are unrelated to rates of mental disorders; (b) in the measurement of mental disorders most PAs have used non-standardized and/or ill-defined instruments in the diagnostic process, and, in cases where standardized instruments have been
used, they have not been validated for use by proxies. Their semi-structured nature also leave them open for interviewer bias; (c) problems related to the informants: emotion-related response bias, no systematic control of type of informants or their relationship with the deceased, number vary by study, the controls are informants themselves in some of the controlled studies, their attitudes toward suicide may influence their responses, and, memory may be affected by how they learned about the suicide; (d) problems related to the interviewers: no control of their psychological and social characteristics, nor how/if they were trained, interviewer perception bias/error; (e) problems related to time between suicide and interview; and (f) problems related to the control groups: matching between cases and controls, and, whether alive or dead controls were used (both have their problems).

These are all serious methodological problems that no doubt affect both the reliability and validity of the findings of these studies. However, usually when the conclusions of a PA study are presented or their results later are used to justify similar studies or in suicide preventive thinking, all the problems seem to be forgotten and only the conclusions remain. For instance, after having pointed to the large number of methodological problems with PA studies, Pouliot and De Leo (2006) concluded, “Regardless of their imperfections, PA studies have contributed substantially to our current knowledge on suicide victims, aetiology of suicide, and suicidal process” (p. 503). However, most of the knowledge produced by PA studies might actually be false if we take all the described weaknesses seriously. This might particularly be the case for the psychiatric diagnoses assigned to the deceased, whereas other types of information may be more reliable.

Cavanagh et al. (2003) in their review took very few (if any) of the weaknesses outlined in previous studies into account. In fact, they explicitly stated that they chose not to use the recommended procedure of calculating a quality score for the studies included in their review but to explore for the effect of possible biases, without explaining why or how this was done. It is also interesting that in the one study included in the review that demonstrated a low prevalence of mental disorder (23%; Rao et al., 1989), the authors dismissed by pointing out that this was a study of self-burning among Indian women. Thus, they did not realize or acknowledge that this actually is a very
important finding in that it may point to cultural differences in the relationship between mental disorder and suicide, which has, in fact, been found (e.g., Chan, Hung, & Yip, 2001; Phillips et al., 2002; Vijayakumar, John, Pirkis, & Whiteford, 2005; Yang et al., 2005; Zhang, Conwell, Zhou, & Jiang, 2004).

**Scrutinizing PA Studies with Particular Focus on the Diagnostic Process**

In the following, we will scrutinize how psychiatric diagnoses are assigned to the deceased in PA studies, followed by a critical discussion of the consequences of recruitment of informants in such studies. We will demonstrate that the methodological weaknesses are so severe that they, in fact, undermine the reliability and hence the validity of psychiatric diagnoses assigned in PA studies.

*Psychiatric Diagnoses Assigned to the Deceased by Interviewing Proxies—Implications for Reliability and Validity*

**INSTRUMENTS USED IN THE DIAGNOSTIC PROCESS**

Pouliot and De Leo (2006) pointed out that many of the PA studies have used unknown or non-standardized diagnostic instruments developed specifically for the respective studies. Often the descriptions of these instruments are vague and thus little is known about their reliability and validity and hence the results of these studies are questionable. A number of studies have, however, used standardized instruments, for instance, different versions of the Schedule for Affective Disorders and Schizophrenia (SADS; Endicott & Spitzer, 1978; Chambers et al., 1985; Orvaschel, Puig-Antich, Chambers, Tabrizi, & Johnson, 1982), the Structured Clinical Interview for various versions of the DSM diagnostic system (SCID; First, Spitzer, Gibbon, Williams, & Benjamin, 1994; First, Spitzer, Gibbon, & Williams, 1995; Spitzer, Williams, Gibbon, & First, 1992), or the Mini International Neuropsychiatric Interview (M.I.N.I.; Sheehan et al., 1992). Several studies have claimed that diagnoses achieved by the use of these instruments in PA studies are reliable and valid (e.g., Brent, Perper, Moritz, Allman, Roth, et al., 1993; Kelly & Mann, 1996; Schneider et al., 2004).
SCRUTINIZING THE CLAIM THAT PSYCHIATRIC DIAGNOSES ASSIGNED IN PA STUDIES ARE VALID

A closer look at some of the studies that have set the tone in terms of validity of diagnoses in PA studies reveals that this claim is made on a somewhat dubious basis. For instance, in their case-control study, Kelly and Mann (1996) compared diagnoses set in PA studies (by means of SCID) with ante-mortem diagnoses set by clinicians that “in most cases” (p. 338) treated the person right before the suicide or another kind of sudden death (controls). They found a high reliability (kappa = 0.84) for Axis I diagnoses and moderate reliability (kappa = 0.65) for Axis II diagnoses. This is the study most often referred to when the validity of diagnoses set in PA studies is asserted. However, one of the inclusion criteria in Kelly and Mann’s study was that the deceased had a medical record of previous psychiatric treatment, and then they interviewed a median of 2 (range 1–4) of the closest relatives. When a close family member has a medical record of psychiatric treatment, it is highly likely that the informants knew about these diagnoses beforehand and thus that their responses are colored by this knowledge. Moreover, only 13 suicides (and 19 controls) were included in this study, certainly a meager number to allow categorical conclusions to be drawn.

Brent, Perper, Moritz, Allman, Roth, et al. (1993) tested the validity of the diagnoses made by proxies by looking at the family history of mental illness (by means of various versions of SADS). Graduate students blind to the original diagnosis interviewed a mean of two first-degree relatives. They found an increased rate of the same disorder in the relatives as the one being assigned to the deceased and claimed that this indicates validity of diagnoses made in PAs. This is hardly a valid argument. Even if there are some genetic factors in mental illness (as certainly have been found; Hamer, 2002), it does not mean that if the mother is depressed, the son will necessarily suffer from depression too. It may be an increased risk, but to use this as a validity check of diagnosis made in PA studies seems a bit far-fetched. Besides, Runeson and Åsberg (2003) found that a family history of suicide predicted suicide independent of mental disorder. And, as Toomela (2007) has pointed out, if the correlation does not equal one, and the unexplained variance is not understood, analyzed, and accounted for, we cannot get a full understanding of the phenomena under study. In other words,
looking at family history of mental illness cannot be a valid assessment of mental illness in a suicide victim.

Brent, Perper, Moritz, Allman, Roth, et al. (1993) also claimed that family history has long been used to validate the diagnosis in a person by referring to an old study only focusing on schizophrenia (Robins & Guze, 1970). Others have also used this reference for claiming the same. One may suspect that these claims have become truisms just because they have been repeated many times. Brent, Perper, Moritz, Allman, Roth, et al. (1993) did, however, list a number of weaknesses of their study: (a) that although the interviewers were blind to the diagnosis, they were not blind to the fact that the person had died by suicide; (b) that high rate of certain disorders in families are related due to diagnostic bias; (c) that a parent suffering from a particular disorder will identify with and report similar symptoms in their children; (d) that a small sample prevented comparison of some diagnostic groups; and (e) that the sample mainly comprised white men. We will add another limitation: that the informants were interviewed and diagnosed by students. Kelly and Mann (1996) emphasized the importance of experienced clinicians in PAs: “The use of...less well trained raters would adversely affect the quality of the psychological autopsy data” (p. 341). Even with all these limitations, Brent, Perper, Moritz, Allman, Roth, et al. (1993) concluded that, “These data provide support for the validity, and therefore, continued use of the psychological autopsy procedure” (p. 121).

The validity of diagnoses assigned by interviewing proxies has also been tested in non-PA studies. For instance, Schneider et al. (2004) interviewed 35 persons randomly selected among the population-based controls for suicides as well as one of these persons’ close relatives (informant) by means of SCID I and SCID II. They found high agreement between the diagnoses obtained from the personal and informant interview. However, the number interviewed is low, and because the controls were from the community, the frequency of diagnoses that could be found was low; so low, in fact, that kappas could not be calculated for some of the disorders most commonly associated with suicide, namely, substance-related disorders, major depression, and schizophrenia. Other weaknesses with this study, as the authors also point to, are that their informants picked out their proxies themselves, and that these proxies had not lost the person in question to suicide; both circumstances
that could have influenced their responses in ways that differ from proxies interviewed following a suicide. In spite of these serious weaknesses, their conclusion is that diagnoses using data from proxies are valid with only minor limitations.

If we look at how Cavanagh et al. (2003) in their comprehensive review of PA studies answer their own question about the reliability of their finding of a strong link between mental disorder and suicide, they start by stating that PAs are considered to be valid in determining diagnoses by referring to Kelly and Mann (1996), the very study we criticized above. Their second argument is that “our results are consistent with the strong positive evidence for the role of mental disorder throughout the suicide literature” (p. 401). However, the evidence they are referring to here is to a large extent based on the very PA studies they reviewed in their present study. Thus, their argument seems a bit tautological. Their third argument is that their study includes all relevant studies and that no publication bias could be detected. However, if all these studies have used the same weak methodology, it doesn’t help that the number of studies is high. Besides, as mentioned earlier they did not assess the studies included in terms of quality. They do, however, point out some potential weaknesses of their study: attribution bias, and “search after meaning” (p. 401), but they do not seem to give much weight to these potential problems because their conclusion is that the best suicide preventive action is treatment of mental disorder.

SCRUTINIZING THE RELIABILITY OF RESPONSES OBTAINED IN PA STUDIES

Regarding reliability, many PA studies have been meticulously concerned with the inter-rater reliability. In some studies inter-rater reliability has been assessed by letting a second person listen to some of the tapes or going through the case notes and/or medical records (e.g., Gustafsson & Jacobsson, 2000). Other studies have had a panel of researchers go through the interviews and/or other records and in this way assessed both inter-rater reliability as well as validity of the diagnoses (e.g., Barraclough, Bunch, Nelson, & Sainsbury, 1974). However, if the responses from the informants are unreliable, inter-rater reliability is worthless. The important issue then is whether the questions in the standardized diagnostic instruments are possible to answer reliably by anyone else than the person being diagnosed. Pouliot and De Leo (2006) have
already pointed out that the instruments used have not been validated for use by proxies, and, that the semi-structured nature leave them open for interviewer bias. To illustrate the serious problems inherent in assigning psychiatric diagnoses to someone by interviewing proxies, we have scrutinized the single items amounting to some of the diagnoses most commonly associated with suicide.

Affective disorders. The diagnoses most often found to be associated with suicide in PA studies are various forms of affective disorders, mainly some forms of depression. Thus, we have looked at the questions from SCID I for the DSM-IV and in the following we present some examples of questions that in our opinion are very hard, if not impossible, to answer reliably by proxies. One of the two main questions to be confirmed as an indication to investigate further for depression is “In the last month, did you lose interest or pleasure in things you usually enjoyed?” If yes, “Was it nearly every day? How long did it last?” If it lasted two weeks or more, one proceeds with follow-up. Can a proxy reliably answer this question? And what about some of the follow-up questions to be answered to reach the diagnostic conclusion about major depressive episode: “How did you feel about yourself?” “Did you have troubles thinking or concentrating?” “Were things so bad that you were thinking a lot about death or that you would be better off dead?” Some of the questions asked to assess melancholic features are “If something good happens to you or someone tries to cheer you up, do you feel better at least for a while?” “Is your feeling of (own equivalent for depressed mood) different from the kind of feeling you would get if someone close to you died?” “Have you been feeling guilty about things you have done or not done?” It must be very difficult for someone else to answer these questions reliably. Take guilt feelings, for instance. Perhaps the informant feels that the deceased should have felt guilty about something and then will respond affirmative here, but it is difficult to know whether the deceased actually felt guilty. Moreover, what if the deceased really had done something bad and felt guilty about it, surely that would have been an appropriate and socially expected reaction, and not a symptom of psychiatric disorder?

Some examples of questions to indicate atypical features that also must be very difficult to answer reliably by proxies are “Do your arms or legs often feel heavy (as though they were full of
“Are you especially sensitive to how others treat you?” “What happens to you when someone rejects, criticizes, or slights you?” “Have you avoided doing things or being with people because you were afraid of being criticized or rejected?” Certainly family members or friends might have opinions about such issues, but do such subjective opinions necessarily represent how the deceased felt?

Some questions from the M.I.N.I. also illustrate the point we are trying to make here. To assign a diagnosis of major depressive episode with this instrument, one of the two main questions to be answered positively in order to justify the follow-up questions are “In the past two weeks, have you been less interested in most things or less able to enjoy the things you used to enjoy most of the time?” Examples of the follow-up questions to establish a diagnosis of major depressive episode are “Did you feel tired or without energy almost every day?” “Did you feel worthless or guilty almost every day?” “Did you have difficulty concentrating or making decisions almost every day?” “Did you repeatedly consider hurting yourself, feel suicidal, or wish that you were dead?” All these questions are very hard to answer reliably on behalf of someone else, even if you feel that you know the person well. And, to complicate matters further, some of these questions overlap with the questions to be answered with regard to generalized anxiety disorder, “Did you feel tired, weak or exhausted easily?” “Did you have difficulty concentrating or find your mind going blank?” Thus, how can one reliably distinguish between these two disorders by this method?

Overlap and co-morbidity. The problem of distinguishing between disorders because of overlapping criteria has also been raised by, for instance, Kelly and Mann (1996; schizophrenia vs. personality disorders), Gunderson and Phillips (1991; borderline personality disorder vs. depressive disorders), and by Brent, Perper, Moritz, Allman, Friend, et al. (1993) who in their study referred to Runeson (1989) and Runeson and Beskow (1991) who had found that a third of young suicides had a borderline personality disorder whereas they themselves found a high frequency of bipolar disorder. However, to explain this discrepancy they stated, “The delineation between cyclothymia, bipolar II, and borderline is difficult to establish, owing to ambiguities in the diagnostic criteria... and it is possible that Runeson and our group are describing essentially the same phenomena” (Brent, Perper, Moritz, Allman,
Friend, et al., 1993, p. 525). This is actually a good argument for the futility of trying to diagnose people by means of interviewing someone else. Not only are the diagnostic criteria ambiguous as Brent, Perper, Moritz, Allman, Friend, et al. (1993) pointed out; studies have found that it is not unusual for people suffering from bipolar disorder to also have a diagnosis of some personality disorder, but this is also very hard to establish even in live subjects because the diagnostic instruments “may not provide the ‘fine tuning’ needed” to establish this (Kutcher, Marton, & Korenblum, 1990). Moreover, it seems difficult to argue that psychiatric disorders are objective and well-defined entities seeing as the DSM system keeps developing both in terms of number of diagnoses included, as well as in terms of definition of the various disorders. For instance, later editions of the DSM have, compared to earlier editions, “made the diagnosis of a major depressive episode more inclusive” (Bostwick & Pankratz, 2000, p. 1925) and thus lowered the threshold for getting this diagnosis. This may very well be another example of the serious methodological problems inherent in PAs that in the end seem to be largely ignored when conclusions are drawn.

If the question of co-morbidity is also taken into consideration, the picture gets even more complicated. Most often some affective disorder is found to be co-morbid with a substance use disorder in PAs (e.g., Brent, Perper, Moritz, Allman, Roth, et al., 1993; Gustafsson & Jacobsson, 2000; Shaffer, 1996). However, if we look at some of the questions asked to establish alcohol abuse/dependence (here exemplified by some of the questions from the M.I.N.I.), we see that they are also difficult to answer reliably by a proxy: “In the past 12 months: Did you need to drink more in order to get the same effect that you got when you first started drinking? During the times when you drank alcohol, did you end up drinking more than you planned when you started? Have you tried to reduce or stop drinking alcohol but failed?” Proxies can speculate, but such speculations can hardly be used to assign psychiatric diagnoses to others.

**Personality disorders.** Personality disorders are also often found in PA studies. However, the questions to be answered to assign such a diagnosis are even harder for others to answer reliably than the ones for Axis I diagnoses described above (e.g., Kelly & Mann, 1996). Besides, studies from other fields have shown low agreement
between self- and informant reports on symptoms for personality disorders (e.g., Klonsky, Olmanns, & Turheimer, 2002; Modestin & Puhan, 2000). Consequently, some authors have not assigned personality disorders in their PA study “because of the difficulty of assessing post mortem the severity, distress, and impairment due to the criteria traits” (Apter et al., 1993, p. 139). In studies including personality disorders, the difficulties result in relatively large groups of a diagnosis of personality disorder not otherwise specified (e.g., Henriksson et al., 1993). Another difficulty in assigning diagnoses of personality disorders is related to the fact that the informant will have to talk negatively about the deceased. On the one hand, this might result in an underestimation of personality disorders because the informant is reluctant to do so, but on the other hand, there might be an overestimation because some informants may have an interest in “blaming” a disease for the suicide and thereby avoid any responsibility. Our point here is not to blame someone specific for a suicide but to stress the fact that people do not live in a vacuum and that we are all affected by our social environment one way or the other.

Consequences of the problems discussed above: Over-/underestimation of diagnoses. From the above, it should be clear that proxies simply cannot answer many of the questions included in the standardized diagnostic instruments reliably. In fact, these questions are exploring the informants’ subjective opinions, feelings, and experiences. Sometimes there might be an underestimation of some diagnosis because the informant is not aware of the things being asked about. However, sometimes there might be an overestimation, for instance, because the informant thinks that you have to be mentally ill to kill yourself, or, because the informant is trying to construct meaning of the suicide and retrospectively interprets behavior in light of this meaning. The study by Gustafsson and Jacobsson (2000) may point toward overestimation. They found that 12 out of 30 outpatients diagnosed in a PA study were actually assigned more diagnoses in the study than what they had received when in treatment just before the suicide. Sometimes informants also might exaggerate symptoms of mental disorder to justify the suicide (Zhang et al., 2004) or to reduce their own feelings of guilt (Wertheimer, 2001).

Another issue that can contribute to an overestimation of psychiatric morbidity in PA studies is the fact that sexual abuse is
a serious risk factor for suicidal behavior (e.g., Evans, Hawton, & Rogham, 2004; Joiner et al., 2006; Martin, Bergen, Richardson, Roeger, & Allison, 2004; Oates, 2004). Often, it is someone in the close family who is the abuser, and if this person happens to be the informant in a PA study, one could hypothesize that it would be convenient for him/her to try to attribute the suicide to mental illness instead of admitting to the abuse. Similar effects may be pertinent to violence.

Zonda (2005) pointed out that another relevant issue is that PA studies do not take the presuicidal syndrome or the crisis situation into consideration and that these conditions cannot be separated from depression in such studies. To separate normal sadness from depression is also difficult, even in screening studies (e.g., Phillips et al., 2007), so it must be even more difficult in PA studies. Moreover, as Gunderson and Phillips (1991) emphasized in their study of the interface between depression and borderline personality disorder, “clinicians should not expect all patients with depressive problems to have an affective illness” (p. 974). Thus, the result could be that depressive disorders are overdiagnosed in PAs (Zonda, 2005).

The main point is, however, that it is difficult or even impossible to know when there might be an overestimation and when there might be an underestimation of psychiatric diagnoses in PA studies. From suicidological studies with other foci than mental disorders, differences have been found between what the person in question and his/her family or doctors say. For instance, Hawton (1982) found large differences between what suicide attempters and their significant others said about the intentionality involved in the suicide attempts. Also, low concordance between children’s self-rating of suicidality compared to the rating of parents and teachers have been found (e.g., Thompson et al., 2006). Why should it be any different in PA studies? Besides, it has been found that the response rate normally is higher from relatives of decedents who were in psychiatric treatment at the time of death, and this can certainly bias the results of PA studies in terms of finding psychiatric diagnoses (Hawton et al., 1998). Moreover, in the instructions to, for example, the M.I.N.I., it is specified that “Clinical judgement by the rater should be used in coding the responses.” Some PA studies have also emphasized that “The investigator’s judgement was required for assigning and excluding diagnoses when the data were incomplete or conflicting” (Henriksson et al., 1993, p. 938). In their review of the methodological concerns of
PA studies, Hawton et al. (1998) actually recommended “making psychiatric diagnoses which seems clinically likely given information about the subject’s behaviour and manner” (p. 273) and are thus not encouraging scientific rigidity. A clinical judgment is bound to be subjective. In the words of Berman (2006): “if I hold to a belief that one has to be mentally disordered to die by suicide, I am more likely to see and report symptoms that fit my belief” (p. 3).

**Inclusion of suicidality as a diagnostic criterion.** For some of the diagnoses, some of the criterion questions are about suicidality. An important issue is thus whether these questions should be included in the diagnostic process or not seeing as the person in question actually has killed him/herself and thus necessarily has fulfilled this criterion at the time of the interview. This makes it easier to reach the required number of questions to be answered positively in order to assign a diagnosis. Hence, controls will less often meet this criterion. Some researchers have explicitly stated that they have excluded this criterion, whereas it is unclear what others have done. If it has been included, this may have artificially raised the proportion of psychiatric disorders compared to controls.

**Number of Informants, Their Mental State, and Their Relationship to the Deceased**

Most PA studies have asked the closest next-of-kin to be informants. This means that for adolescent suicides, most often parents have been interviewed, sometimes with the addition or replacement of siblings and/or friends. In studies of adult suicides, the most common informant has been a spouse or other close family members like children, or siblings. However, it is not necessarily the closest in kin who is the closest one in terms of confidence or intimacy. Some studies have found that, for instance, friends sometimes know more than parents about substance abuse, interpersonal problems, and/or suicidal ideation (e.g., Hawton et al., 1998). Thus, different informants can contribute with different information and this makes it crucial to know “who” we are interviewing in terms of intimacy of the relationship between the informant and the deceased. For instance, our research group is currently conducting a qualitative PA study where we interview in depth at least five persons around each suicide. Preliminary analyses show that the stories vary
considerably from one informant to the next. This underlines the importance the relationship between the informant and the deceased has for the information given. However, very few, if any, studies have discussed, let alone mentioned, the importance of the character of the relationship between the suicide victim and the informants and its significance for the information obtained.

Many PA studies have only interviewed one or two close persons around each suicide and it is impossible from the descriptions in the studies to know who these persons are because closeness is poorly defined. Hardly any of the PA studies discuss the issue about closeness in relationship between deceased and informant. Some studies have, however, interviewed more than one or two informants around each suicide and thus increased the chance of including real knowledgeable ones. For instance, Apter et al. (1993) interviewed an average of 10 informants for each suicide; Barraclough et al. (1974) interviewed an average of 4.5 per suicide; Brent et al. (1991; Brent, Perper, Moritz, Allman, Roth, et al., 1993; Brent, Johnson, Perper, et al., 1994a; Brent, Perper, Moritz, et al., 1994b) interviewed a median of three to four informants (range 1–14 in the different studies); Conwell et al. (2000) interviewed a mean of 4.3 informants; and Runeson (1989) and Runeson and Beskow (1991) interviewed a mean of 2.2 (range 1–5) per suicide. However, in most of these studies it is unclear whether the different informants are interviewed separately or together; whether they have had the chance to discuss the interview with each other if they were interviewed separately; or how a diagnosis was reached when several informants were interviewed (by adding symptoms from the different sources?). Often it is stated that diagnoses were assigned “by a ‘best estimate’ consensus process” (e.g., Conwell et al., 2000, p. 28), or “best estimate, based on all available data” (Runeson & Beskow, 1991, p. 153).

In addition, very few of the PA studies have described how they have dealt with disagreements between sources where more than one have been used, even though this is one of the 12 standards for PA studies outlined by Clark and Horton-Deutsch (1992). And if the issue is raised, normally there is no discussion of the consequences of the way they handled it. For instance, Palacio et al. (2007) interviewed two relatives to begin with and if there were disagreements between them, a third person was interviewed. However, how can we be sure that the majority has the true version of
the deceased’s story? Brent et al. (1991; Brent, Perper, Moritz, Allman, Roth, et al., 1993) handled disagreement between informants by re-interviewing them “around the areas of discrepancies until satisfactory consensus could be obtained” (1993, p. 522). There is no description of how consensus was obtained or how satisfactory consensus was defined. Sometimes the research team has just made a decision after reviewing the case in detail (Yang et al., 2005), and sometimes the higher value was chosen for analysis which certainly may have led to an overestimation of psychiatric symptoms (Zhang et al., 2004).

The mental state of the informant is also an important factor in this discussion (Pouliot & De Leo, 2006). Research from other fields have shown that depressed/distressed mothers report more affective symptoms in their children than the children themselves report (Garber, Van Slyke, & Walker, 1998), and that informants with emotional problems tend to remember more negative information (Moradi, 2000). If the bereaved are interviewed soon after the suicide, which is the case in some of the PA studies (e.g., an extreme is the study of Apter et al., 1993, where the interviews took place only a week after the suicide), they might still be in shock. Many will probably show a number of depressive symptoms and their responses will be colored by this state of mind. On the other hand, if it is a very long time since the suicide, important details may have been forgotten and it will be very hard to respond to questions about the last two weeks of the life of the deceased (which is necessary to establish some of the diagnoses). There are also a number of other problems connected to the informants in PA studies (Pouliot & De Leo, 2006), but they are beyond the scope of the present article to discuss.

Variations Between Studies

There are large variations from one PA study to the next in terms of rates of the different types of mental disorders in PAs (e.g., Bertolote, Fleischman, De Leo, & Wasserman, 2004; Cavanagh et al., 2003). Several factors can explain these differences, for instance; that the samples included varied in terms of where they were recruited from (e.g, general population, psychiatric hospitals); that there are age, gender, regional, and/or cultural differences in the diagnostic pattern of mental disorders; or, simply that the PA method is not reliable and/or valid so that the results are random
(see our discussion above). Even when PA studies with a case-control design find much larger proportions of suicides connected to mental disorder compared to controls, the size of the difference might have been blown out of proportion because of the expectations of mental disorder connected to suicide in both informants and interviewers (Berman, 2006; Pouliot & De Leo, 2006). The result in case-control studies is also to some degree dependent upon whether the diagnostic criterion of suicidality is included or not.

**Conclusions and Implications for Future Research**

To answer the question raised in the title of this article explicitly: Yes, as a tool to assign psychiatric diagnoses to dead people by interviewing proxies, PA studies are methodologically flawed. It is simply impossible to assign a reliable diagnosis of mental disorder to someone by interviewing someone else. PA studies can therefore not serve as an evidence base for the claim that most people who die by suicide are mentally ill. Although many PA studies have used other sources of information in addition to interviews of bereaved (e.g., medical/psychiatric records, files from general practitioners, police records, coroners’ reports, etc.), there are hardly any reports on how such information of the different sources is handled. If the information is just summarized, that is, if some of the diagnostic criteria are found in journals, some from the interview of one informant and others from other informants, this can hardly be considered a valid way of assigning diagnoses. To compare diagnoses from interviews with diagnoses found in psychiatric records has also proven difficult (Bertolote et al., 2003; Gustafsson & Jacobsson, 2000). We have, however, not disproven that there might be a relationship between mental disorders and suicide; we simply argue that the results of PA studies do not constitute a valid evidence base for such a relationship.

Pouliot and De Leo (2006) concluded their review of all the methodological problems inherent in PA studies by recommending more standardization of the method. Our conclusion is in some way the opposite: It is not more standardization that is needed, but less, or perhaps rather, more adjusted systematism. That is, we need qualitative approaches focusing on the understanding of suicide beyond mental disorders, which is also more in line with the original purpose of the method (Shneidman, 1981). No matter how
standardized the diagnostic manuals become, we still have the problem with reliability of the information from proxies. Thus, the common way of conducting PAs with the main focus on psychiatric diagnoses should now be abandoned. Instead, we encourage qualitative interviews of a relatively high number of informants around each suicide where each of them is given the opportunity to tell their story of why their loved one killed him-/herself. Then, we recommend that these narratives be systematically analyzed in terms of the informants’ relationships with the deceased (not just in terms of kin). Such studies will make it more possible to reveal why that particular person at that particular time decided to end his/her life. This will, in turn, contribute to our overall understanding of what suicide is all about, what it means to the suicidal persons themselves, in their particular contexts, and thus contribute to inform practice in suicide prevention.

References

Apter, A., Bleich, A., King, R. A., Kron, S., Fluch, A., Kotler, M., & Cohen, D. J. (1993). Death without warning? A clinical post-mortem study of suicide in 43 Israeli adolescent males. *Archives of General Psychiatry, 50*, 138–142.
Barracough, B., Bunch, J., Nelson, B., & Sainsbury, P. (1974). A hundred cases of suicide: Clinical aspects. *British Journal of Psychiatry, 125*, 355–373.
Berman, A. (2006). The other 10 percent. *Newslink*, 33(3), 3.
Bertolote, J. M., Fleischman, A., De Leo, D., & Wasserman, D. (2003). Suicide and mental disorders: Do we know enough. *British Journal of Psychiatry, 183*, 382–383.
Bertolote, J. M., Fleischman, A., De Leo, D., & Wasserman, D. (2004). Psychiatric diagnoses and suicide: Revisiting the evidence. *Crisis, 25*, 147–155.
Beskow, J., Runeson, B., & Åsgård, U. (1990). Psychological autopsies: Methods and ethics. *Suicide and Life-Threatening Behavior, 20*, 307–323.
Bostwick, J. M., & Pankratz, V. S. (2000). Affective disorders and suicide: A re-examination. *American Journal of Psychiatry, 157*, 1925–1932.
Brent, D. A., Johnson, B. A., Perper, J., Conolly, J., Bridge, J., Bartle, S., & Rather, C. (1994). Personality disorder, personality traits, impulsive violence, and completed suicide in adolescents. *Journal of American Academy of Child and Adolescent Psychiatry, 33*, 1080–1086.
Brent, D. A., Perper, J. A., Allman, C. J., Moritz, G. M., Wartella, M. E., & Zelenak, J. P. (1991). The presence and accessibility of firearms in the homes of adolescent suicides. *JAMA, 266*, 2989–2995.
Brent, D. A., Perper, J. A., Moritz, G., Allman, C., Friend, A., Roth, C., et al. (1993). Psychiatric risk factors for adolescent suicide: A case-control study. *Journal of American Academy of Child and Adolescent Psychiatry, 32*, 521–529.
Brent, D. A., Perper, J. A., Moritz, G., Allman, C. J., Roth, C., Schweers, J., & Balach, L. (1993). The validity of diagnoses obtained through the psychological autopsy procedure in adolescent suicide victims: Use of family history. *Acta Psychiatrica Scandinavica, 87*, 118–122.

Brent, D. A., Perper, J. A., Moritz, G., Baugher, M., Schweers, J., & Roth, C. (1994). Suicide in affectively ill adolescents: A case-control study. *Journal of Affective Disorders, 21*, 193–202.

Cavanagh, J. T. O., Carson, A. J., Sharpe, M., & Lawrie, S. M. (2003). Psychological autopsy studies of suicide: a systematic review. *Psychological Medicine, 33*, 395–405.

Chambers, W. J., Puig-Antich, J., Hirsch, M., Paez, P., Ambrosini, P. J., & Tabrizi, M. A. (1985). The assessment of affective disorders in children and adolescents by semi-structured interview: Test-retest reliability of the schedule for affective disorders and schizophrenia for school-aged children, present episode version. *Archives of General Psychiatry, 42*, 696–702.

Chan, K. P. M., Hung, S. F., & Yip, P. S. F. (2001). Suicide in response to changing societies. *Child and Adolescent Psychiatric Clinics of North America, 10*, 777–795.

Clark, D. C., & Horton-Deutch, S. L. (1992). Assessment in absentia: The value of the psychological autopsy method for studying antecedents of suicide and predicting future suicides. In R. W. Maris, A. L. Berman, J. T. Maltsberger, & R. I. Yufit (Eds.), *Assessment and prediction of suicide* (pp. 144–182). New York, NY: The Guilford Press.

Conwell, Y., Lyness, J. M., Duberstein, P., Cox, C., Seidlitz, L., Digiorgio, A., & Caine, E. D. (2000). Completed suicide among older patients in primary care practices: A controlled study. *Journal of the American Geriatrics Society, 48*, 23–29.

Endicott, J., & Spitzer, R. L. (1978). A diagnostic interview: The schedule for affective disorders and schizophrenia. *Archives of General Psychiatry, 35*, 837–844.

Evans, E., Hawton, K., & Rogham, K. (2004). Factors associated with suicidal phenomena in adolescents: A systematic review of population-based studies. *Clinical Psychology Review, 24*, 957–979.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1995). *Structured clinical interview for DSM-IV Axis I disorders*. New York, NY: Biometrics Research Department.

First, M. B., Spitzer, R. L., Gibbon, M., Williams, J. B. W., & Benjamin, L. (1994). *Structured clinical interview for DSM-IV Axis II personality disorders*. New York, NY: Biometrics Research Department.

Garber, J., Van Slyke, D. A., & Walker, L. S. (1998). Concordance between mothers’ and children’s reports of somatic and emotional symptoms in patients with recurrent abdominal pain or emotional disorders. *Journal of Abnormal Child Psychology, 26*, 381–391.

Gunderson, J. G., & Phillips, K. A. (1991). A current view of the interface between borderline personality disorder and depression. *American Journal of Psychiatry, 148*, 967–975.
Gustafsson, L., & Jacobsson, L. (2000). On mental disorder and somatic disease in suicide: A psychological autopsy study of 100 suicides in northern Sweden. *Nordic Journal of Psychiatry, 54*, 383–395.

Hamer, D. (2002). Rethinking behaviour genetics. *Science, 298*, 71–72.

Hawton, K. (1982). How patients and psychiatrists account for overdoses. In E. Shepherd & J. P. Watson (Eds.), *Personal meanings* (pp. 103–114). New York, NY: Wiley.

Hawton, K., Appleby, L., Platt, S., Foster, T., Cooper, J., Malmberg, A., & Simkin, S. (1998). The psychological autopsy approach to studying suicide: A review of methodological issues. *Journal of Affective Disorder, 50*, 269–276.

Henriksson, M. M., Aro, H. M., Marttunen, M. J., Heikkinen, M. E., Isometsä, E. T., Kuoppasalmi, K. I., & Lönnqvist, J. K. (1993). Mental disorders and comorbidity in suicide. *American Journal of Psychiatry, 150*, 935–940.

Isacsson, G., & Rich, C. L. (2003). Getting closer to suicide prevention. *British Journal of Psychiatry, 182*, 455–459.

Joiner, T. E., Sachs-Ericsson, N. J., Wingate, L. R., Brown, J. S., Anestis, M. D., & Selby, E. A. (2007). Childhood physical and sexual abuse and lifetime number of suicide attempts: A persistent and theoretically important relationship. *Behaviour Research and Therapy, 45*, 539–547.

Kelly, T. M., & Mann, J. J. (1996). Validity of DSM-III–R diagnosis by psychological autopsy: A comparison with clinician ante-mortem diagnosis. *Acta Psychiatraca Scandinavica, 94*, 337–343.

Klonsky, E. D., Oltmanns, T. F., & Turheimer, E. (2002). Informant-reports of personality disorder: Relation to self-reports and future research directions. *Clinical Psychology: Science and Practice, 9*, 300–311.

Kutcher, S. P., Marton, P., & Korenblum, M. (1990). Adolescent bipolar illness and personality disorder. *Journal of American Academy of Child and Adolescent Psychiatry, 29*, 355–358.

Martin, G., Bergen, H. A., Richardson, A. S., Roeger, L., & Allison, S. (2004). Sexual abuse and suicidality: Gender differences in a large community sample of adolescents. *Child Abuse & Neglect, 28*, 491–503.

Modestin, J., & Puhan, A. (2000). Comparison of assessment of personality disorder by patients and informants. *Psychopathology, 33*, 265–270.

Moradi, A. R., Taghavi, R., Neshat-Doost, H. T., Yule, W., & Dalgleish, T. (2000). Memory bias for emotional information in children and adolescents with posttraumatic stress disorder: a preliminary study. *Journal of Anxiety Disorders, 14*, 521–534.

Mortensen, P. B., Agerbo, E., Erikson, T., Qin, P., & Westergaard-Nielsen, N. (2000). Psychiatric illness and risk factors for suicide in Denmark. *The Lancet, 355*, 9–12.

Oates, K. R. (2004). Sexual abuse and suicidal behaviour. *Child Abuse & Neglect, 28*, 487–489.

Orvaschel, H., Puig-Antich, P., Chambers, W., Tabrizi, M. A., & Johnson, R. (1982). Retrospective assessment of prepubertal major depression with the Kiddie-SADS-E. *Journal of American Academy of Child and Adolescent Psychiatry, 27*, 392–397.
Palacio, C., Garcia, J., Diago, J., Zapata, C., Lopez, G., Ortiz, J., & Lopez, M. (2007). Identification of suicide risk factors in Medellin, Colombia: A case-control study of psychological autopsy in a developing country. Archives of Suicide Research, 11, 297–308.

Phillips, M. R., Shen, Q., Liu, X., Pritzker, S., Streiner, D., Conner, K., & Yang, G. (2007). Assessing depressive symptoms in persons who die of suicide in mainland China. Journal of Affective Disorders, 98, 73–82.

Phillips, M. R., Yang, G., Zhang, Y., Wang, L., Ji, H., & Zhou, M. (2002). Risk factors for suicide in China: A national case-control psychological autopsy study. The Lancet, 360, 1728–1736.

Portzky, G., Audenhart, K., & Van Heeringen, K. (2005). Suicide among adolescents. A psychological autopsy study of psychiatric, psychosocial and personality-related risk factors. Social Psychiatry and Psychiatric Epidemiology, 40, 922–930.

Pouliot, L., & De Leo, D. (2006). Critical issues in psychological autopsy studies. Suicide and Life-Threatening Behavior, 36, 491–510.

Rao, A. V., Mahendran, N., Gopalakrishnan, C., Reddy, T. K., Prabhakar, E. R., Swaminatan, R., et al. (1989). One hundred female burn cases: A study in suicidology. Indian Journal of Psychiatry, 7, 330–333.

Robins, E., & Guze, S. B. (1970). Establishment of diagnostic validity in psychiatric illness: Its application to schizophrenia. American Journal of Psychiatry, 126, 983–987.

Runeson, B. (1989). Mental disorder in youth suicide. DSM-III–R Axes I and II. Acta Psychiatr. Scandinavica, 79, 490–497.

Runeson, B., & Åsberg, M. (2003). Family history of suicide among suicide victims. American Journal of Psychiatry, 160, 1525–1526.

Runeson, B., & Beskow, J. (1991). Borderline personality disorder in young Swedish suicides. The Journal of Nervous and Mental Disease, 179, 153–156.

Schneider, B., Maurer, K., Sargk, D., Heiskal, H., Weber, B., Frölich, L., et al. (2004). Concordance of DSM-IV Axis I and II diagnoses by personal and informant’s interview. Psychiatry Research, 127, 121–136.

Shaffer, D., Gould, M. S., Fisher, P., Trautman, P., Moreau, D., Kleinman, M., & Flory, M. (1996). Psychiatric diagnosis in child and adolescent suicide. Archives of General Psychiatry, 53, 339–348.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., et al. (1992). The Mini international neuropsychiatric interview (M.I.N.I.): The development and validation of a structured diagnostic interview for DSM-IV and ICD-10. Journal of Clinical Psychiatry, 20(Suppl.), 34–57.

Shneidman, E. (1981). The psychological autopsy. Suicide and Life-Threatening Behavior, 11, 325–340.

Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. B. (1992). The structured clinical interview for DSM-III–R (SCID) I: History, rationale, and description. Archives of General Psychiatry, 49, 624–629.

Thompson, R., Dubowitz, H., English, D. J., Noonan, K. B., Wike, T., Bangdiwala, S. I., et al. (2006). Parents’ and teachers’ concordance with children’s self-ratings of suicidality: Findings from a high-risk sample. Suicide and Life-Threatening Behavior, 36, 167–181.
Toomela, A. (2007). Culture of science: Strange history of the methodological thinking in psychology. *Integrative Psychological and Behavioral Science, 41*, 6–20.

Vijayakumar, L., John, S., Pirkis, P., & Whiteford, H. (2005). Suicide in developing countries: 2. Risk factors. *Crisis, 26*, 112–119.

Wertheimer, A. (2001). *A special scar. The experiences of people bereaved by suicide* (2nd ed.). London, UK: Routledge.

Yang, G.-H., Phillips, M. R., Zhou, M. G., Wang, L.-J., Zhang, Y.-P., & Xu, D. (2005). Understanding the unique characteristics of suicide in China: National psychological autopsy study. *Biomedical and Environmental Sciences, 18*, 379–389.

Zhang, J., Conwell, Y., Zhou, L., & Jiang, C. (2004). Culture, risk factors and suicide in rural China: A psychological autopsy case control study. *Acta Psychiatrica Scandinavica, 110*, 430–437.

Zonda, T. (2005). Depression and suicidal behaviour. *Crisis, 26*, 34–35.