

Making Clinical Research Empirically Relevant: Drew Westen's lecture at the Sapienza (14 June 2006)¹

by Francesco Gazzillo² and Vittorio Lingiardi³

The issue of therapies based on empirical evidence

The epistemological objections to psychoanalytic theory put forward by some scholars (Grünbaum, 1984), the doubts of some researchers on the real efficacy of the psychotherapies (Eysenck, 1952), the development of pharmacological therapy effective for psychiatric disorders, and the requests of the insurance companies, willing to finance treatments whose efficacy is short term and empirically proven, are some of the factors that triggered a research movement, mainly with cognitive-behavioural roots, aimed at elaborating standardised psychotherapy protocols for specific disorders, with experimentally proven efficacy.

This was how the movement of *Empirically Supported Treatments* (EST) (Kendall, 1998) originated, also driven by the desire to avoid an imbalance present in the mental health market, populated by the most variegated forms of therapy but lacking data about the relation between the problem for which a therapy is sought, the costs, duration and outcome of the treatment.

At a time when to be marketable, a drug or a therapy must prove its efficacy in carefully controlled studies, also psychotherapy – considered a medical therapy – must be promoted on the market of rival therapies and before possible users with a pedigree of empirical proof as to its efficacy.

The needs of the experimentation on EST

The requirements of research and the kind of epistemology and therapeutic modelling referred to by the EST researchers place specific constraints on the studies aimed at elaborating and assessing treatments (Chambless & Ollendick, 2000; Kendall, Marrs-Garcia, Nath, & Shedrick, 1999; Nathan, Stuart, & Dollan, 2000). Here are some of them:

1. psychotherapy validation studies must exclude subjects that undermine the homogeneity of the experimental and control samples;
2. the treatments to assess must be short, or at any rate must have a pre-established duration, because the protocol must be the same in all cases and the clinical situation presented by the patients at the beginning, must at the end of the therapy and in the follow-up be just as assessable as all the other conditions.
3. the treatments must be aimed at the treatment of a single clinical disorder that can be diagnosed with the psychopathological guidelines accepted by the international community (Goldfried, 2000) and they must be specific to that illness. This is why from the EST validation samples those patients who present comorbidity should be eliminated and the state and possible changes presented by disorders different from those targeted should be ignored. (Wilson, 1998);
4. the treatments to assess must be manual-based and the degree of adherence to the protocol of the real practice of the professionals involved in the studies must be monitored.

¹ In this article we have summed up, with the agreement of Drew Westen, the main topics of his paper "Making Clinical Research Empirically Relevant", held in Rome on 14 June 2006 in the Aula Magna of the Faculty of Psychology (Introduction by Nino Dazzi and Vittorio Lingiardi, consecutive translation by Francesco Gazzillo).

² PhD Student in Clinical, Dynamic and Development Psychology, Faculty of Psychology 1, University "La Sapienza", Rome.

³ Special Professor, Department of Clinical, Dynamic and Development Psychology, Faculty of Psychology 1, University "La Sapienza", Rome.

In short, the EST movement tries to elaborate brief psychotherapy treatments, manual-based and aimed at the treatment of specific disorders, and it assesses them by means of samples of homogeneous subjects treated (and/or assessed) in an identical way and for the same target pathology.

EST and the needs of real clinical practice

To judge the clinical relevance and the ecological validity of EST, it is necessary to answer a series of questions first.

1. What problems usually bring people to begin psychotherapy initially? Are these problems the ones EST aim to solve? If not, it is as if the EST movement were trying to find reliable answers to questions posed rarely or by very few people.
2. Are the treatments whose efficacy we assess the same as the ones that are carried out in real clinical practice? If not, it is as if the EST were doing empirically sound research into the running of the unicorns.
3. What is the optimum duration of a psychotherapy in relation to the disorder that it is trying to cure and to the outcomes it manages to obtain? Is there a dose-effect relation in the psychotherapy field?
4. What is the relation between the results obtained in the research into EST and those of the therapies actually carried out in public and private clinical practice?
5. On the basis of what criteria do we claim that a certain kind of psychotherapy has empirical support and another does not? Do these criteria, besides respecting the needs of empirical research, have any clinical relevance at all?⁴
6. How can we integrate the results of the EST research and the discoveries of general psychology and neuroscience?

And the first big problem we encounter, argues Westen, is that all these questions have not yet received an answer.

The problems of those who enter therapy

Concerning the type of psychic pathologies for which patients seek treatment, most of the research indicates that in general psychotherapy is not begun for a single disorder that can be diagnosed in the DSM, and that around half the patients with a disorder that can be diagnosed in Axis II also present Axis I disorders and viceversa (Newman, Moffit, Caspi, & Silva, 1998; Zimmermann, McDermut, & Mattia, 2000). Many patients present below-threshold Axis I disorders, i.e. symptoms of the higher disorders which however do not satisfy the algorithms established by the DSM for a diagnosis of disorder; moreover, most patients who seek a therapy present personality disorders falling below threshold (Howard, et al., 1996; Messer, 2001). In all cases, there are no EST studies related to below-threshold pathologies.

Lastly, we have no empirical proof in support of the hypothesis that different people diagnosed with the same disorder can benefit to the same extent from the same kind of treatment. And it is hardly plausible, for example, that a young man with depressive syndrome, confusion on his sexual orientation and serious family conflicts can benefit in

⁴At this point it is useful to recall the difference between *efficacy* and *effectiveness*. *Efficacy research* refers to laboratory-controlled studies (for instance, a study on psychotherapy that monitors, controls, and standardizes the treatment procedures and uses randomized groups that are homogeneous with each other). Instead we talk about *effectiveness research* when the experimenters highlight the fact that the treatment is carried out in a clinical context, perhaps less controlled than in a laboratory, but more valid from the ecological point of view. For Kazdin (2002), *efficacy* and *effectiveness* should be considered extreme poles of a continuum rather than discrete categories, and the researcher's task is to find a good equilibrium between the two (knowing that a sacrifice of *efficacy* usually favours the clinical relevance of the results).

the same way and from the same treatment that helps a middle-aged woman with major depressive disorder, who has recently entered menopause. The DSM diagnosis in both cases could be single major depressive episode, but the conditions of the subjects seem rather different. It is hard to believe that the same psychotherapy – manual-based, what is more – has the power to alleviate the depression of these two subjects in the same way. In short, we have no empirical proof to support the clinical relevance of research which – like that on EST – starts from the premise that the psychotherapies must be effective for the single disorders diagnosed in the DSM and with no comorbidity. At the same time there is proof to support the idea that patients that seek psychotherapy do not do so because they suffer from one single disorder that can be diagnosed with the criteria of the DSM.

The problem of the type of treatment to assess

The data available on the choice of treatments to assess empirically is also rather discouraging. While it would be better for the effectiveness of the most commonly used treatments among the community of real patients and clinicians to be assessed empirically, what has in fact been assessed so far by most of the EST research is almost exclusively the efficacy of short term cognitive-behavioural and interpersonal therapies. The assumption of this type of study is that it is possible to modify a psychic pathology in a short, pre-established period of time. But also in this case most of the research available shows how difficult it is to have an effect on psychic pathologies, above all in a limited time-span; and the high level of relapses suffered by people undergoing psychotherapy indicates the uncertain nature of many of these results.

Moreover, while research on EST tries to avoid the problem of comorbidity of Axis I and Axis II disorders by excluding from the experimental samples subjects with personality disorders, and not monitoring the possible changes in these disorders during therapy, empirical research indicates unequivocally that personological conditions constitute the diathesis of Axis I disorders (Westen, Gabbard, & Blagov, 2006), but the needs of the EST experimentation do not allow research that takes this fact into account; comorbidity is considered merely a juxtaposition of disorders, and for each disorder there should be a different treatment. For Axis II disorders, however, there are few EST treatments available.

The duration of treatments

There is no data to enable us to make a hypothesis on the optimum duration of a psychotherapy treatment. While the treatments empirically assessed generally last 4-20 sessions, comparable to the duration of drug effectiveness controls (Goldfried, 2000), research indicates that treatments lasting 24 months or more are generally more effective than short ones, especially when there are Axis II problems (Howard, Kopta, Krauser, & Orlinsky, 1986; Kopta, Howard, Lowry, & Beutler, 1994; Seligman, 1995). The high level of relapses suffered by subjects treated with EST is in line with this fact.

The optimum alternative to the EST: what does the control group do?

Another important issue is that of the relation between the results obtained with the EST and those obtained from skilled psychotherapists working in the community. On this point, too, the information at our disposal does not allow us to draw reliable conclusions; this is due firstly to the fact that, in general, the control samples used in experiments for empirical assessment of the efficacy of the EST are from support psychotherapies which exclude the possibility of discussing symptoms, or are waiting list patients. Both cases constitute artificial conditions, outside the competence of any mental health professional and are explicitly non-therapies. Given these premises, the EST results could be explained not by seeking to demonstrate their efficacy but in a different way, i.e. starting from the assumption that a therapy that wants to have an effect on the psychic conditions of a patient is more effective than one that does not. EST experiments would gain greater

clinical relevance by using control groups of patients in therapy with skilled professionals working in the way that suits them best.

The efficacy of a treatment according to the EST movement

A treatment is generally considered validated when it has been subjected to a certain amount of empirical research, the results of which, assessed with *meta-analysis*⁵, show a positive change compared to the clinical conditions found at the beginning of the therapy and/or in the control group.

The *effect size*⁶ of this change must be adequate; the percentage of patients who at the end of the therapy have improved or been cured must be higher than those of patients who have given up the treatment or who have received no benefit; the symptoms present at the beginning of the treatment must be reduced or absent at the end and the results of the studies must be generalizable. Lastly, the percentage of patients who have relapses or seek other treatment after the conclusion of the treatment, must be taken into account.

Some figures on three examples of Empirically Supported Treatments

To assess to what extent these requisites are satisfied by the studies carried out so far on EST, we can consider two broad research contexts: manual-based treatments for panic attacks and those for depression (Westen & Morrison, 2001).

On the former, the percentage of patients with panic attacks included in the samples for the validation of these therapies is about 36%; 86% of these people completed the therapy and 63% of the patients that finished the treatment did not present further symptoms. After the therapy, the patients presented on average 2 panic attacks every 4 weeks (0,7 panic attacks a week with a standard deviation of 1.2) but two years after the conclusion of the therapy, only 46,4% of the patients had maintained the improvement unaltered, while 49% of the patients presented relapses and/or sought another therapy. In other words, out of 100 people with panic attacks, only 20 patients were able to use an empirically supported treatment and at the end of the therapy no longer presented symptoms, and of these only 9 maintained this improvement in the two years following the conclusion of the therapy. As far as the manual-based therapies for the treatment of depression are concerned, the data tells us that only 36% of the depressed patients satisfied the criteria for inclusion in these studies; of them, 74% completed the therapy. 54% of the latter showed an improvement. We do not know how many of these cases no longer showed depressive symptoms at the end of the therapy, but 12-18 months after the conclusion of the treatment, only 29% of the patients did not present relapses, while around 54% sought another therapy in the two years following the end of the first treatment. Out of 100 depressives, therefore, 15 completed an EST and at the end showed improvements, but only 4 maintained these improvements in the following one or two years.

One last case is that of the treatments empirically validated for post-traumatic stress disorder. In March 2005 the *National Institute for Clinical Excellence* of the USA laid down the guidelines for the treatment of this disorder, and according to these guidelines the only therapies that have been shown to be valid for PTSD are cognitive-behavioural therapy and *Eye Movement Desensitization Retraining* (EMDR). Also in this case, however, the generalizability of the results is in doubt, above all because of the exclusion of polysymptomatic patients and of patients presenting personality disorders or histories of serious traumas, especially in childhood – all conditions that are frequently associated with the presence of PTSD in adulthood. Lastly, rather a high percentage of patients treated

⁵ Meta-analysis applied to research on EST is a procedure that makes it possible to assess quantitatively the effectiveness of a certain treatment on the basis of the results of studies conducted in different places, times and conditions.

⁶ The *effect size* is a measurement of the difference between the average conditions of symptoms presented by patients in the experimental group at the end of the therapy and the average conditions of the control group (i.e. without therapy) compared to the standard deviation of the sample as unit of measurement and supposing a normal distribution of this data.

with EST for PTSD presented residual symptoms at the end of the therapy, the follow-up data concerned rather limited periods of time and we have no empirical data on the levels of improvement and cure of patients treated with non-manual-based psychotherapies conducted by skilled clinicians.

Summing up, the research of the EST movement provided us with enough empirical proof to sustain an attitude of extreme *humility*: the improvements due to psychotherapy are in general only partial, the progression of the clinical conditions seesaws in the long term and after the end of the psychotherapy most of the patients present relapses and seek further therapy. Above all, as far as most of the questions on the therapy of psychic disorders are concerned, the research has not yet managed to provide empirically sound, clinically relevant answers.

Some of Westen's proposals: the empirically informed psychotherapies

To get out of the present situation – which according to Westen can be summed up in the idea that many professional mental health workers, in order to carry out controlled experimentation on the efficacy of psychotherapy, have rushed into a sort of “uncontrolled mass experimentation” – it would be wise to move from the idea of empirically grounded therapies to that of *empirically informed psychotherapy*, a way of not seeking easy answers to complex questions.

Here are some pointers provided by Westen to carry out this ambitious project:

1. it would be a good idea to let clinicians do what they know best, i.e. observe, listen, infer and describe, leaving to researchers and statisticians the task within their competence, i.e. collect and analyse data in the most correct and informative way possible (Westen & Weinberger, 2004);
2. clinicians and researchers with different origins, training and theoretical orientation should be involved in multifocal experimental research by means of a working network in which the clinicians collect the information and the researchers elaborate it (Westen, Shedler, & Bradley, in press);
3. as well as assessing the dimensions of directly observable or self-described psychic functioning, research must translate the main inferential constructs of the psychological and psychotherapy field into their objective correlates, and use descriptors that do not resort to jargon belonging to specific theoretical approaches to assess them at an intermediate level of abstraction.

These are the strategies adopted by Westen and collaborators in the construction of tools to assess constructs like that of personality (Westen & Shedler, 1999 a, b), object relations and social cognitions (Westen, 1990), transference (Bradley, Heim, & Westen, 2005), countertransference (Betan, Heim, Concklin, & Westen, 2005), attachment (Nakash-Eisikovits, Dutra, & Westen, 2002), identity disorder (Wilkinson-Ryan, Westen, 2000), affective regulation (Zittel, Bradley, & Westen, 2006).

SWAP and its applications

The *Shedler-Westen Assessment Procedure-200* (SWAP-200) (Westen, Shedler, & Lingardi, 2003) – of which there is also a version for adolescents (SWAP-200-A) and a second version being validated both for adults and for adolescents (SWAP-II e SWAP-II-A) – is the implementation of a Q-sort procedure in personality assessment.

It consists of 200 statements describing normal and pathological personality traits. These statements were elaborated on the basis of the literature on personality and personality disorders, defence and coping mechanisms; of research into personality and its disorders; of Axis I and II and the appendixes of editions III-R and IV of the *Diagnostic and Statistical Manual of Mental Disorders* (APA 1987; APA 1994). The original list of SWAP descriptors was progressively modified and refined on the basis of its early applications, eliminating

the redundant items, those with too low a variance etc. All the items of SWAP-200 are written in simple language without technical terms, and try to investigate both directly observable aspects of the personality, like the tendency to overeat followed by purges (item 70) or to react to criticisms with feelings of anger and humiliation (item 103), and hypothetical constructs like that of projective identification or reactive formation. Where the objects of study are inferential constructs, however, the authors have broken them down into their directly observable phenomenal correlates and devoted at least one item to the description of these correlates. For instance, the concept of projective identification is studied by means of three items: 116 "Tends to see her/his unacceptable feelings and impulses in others and not in her/himself" (projective component); 76: "Behaves so as to arouse in others feelings similar to those s/he is experiencing (e.g., when s/he is angry, s/he acts in such a way that others are provoked to anger; when s/he is anxious, s/he acts in such a way as to arouse anxiety in others)" (interpersonal manipulation); 154: "Tends to arouse extreme reactions and strong feelings in others" (consequences of the previous operations on other people). In this sense, SWAP-200 tries to use the clinical experience accumulated in over 100 years of psychodynamic thought without sacrificing the needs of scientific research in the strict sense.

The SWAP-200 items describe traits related to the 4 dimensions that have to be explored to obtain a *functional assessment of personality* (Westen, 1998) in order to bridge the gap between descriptive diagnosis and case formulation, and sees personality traits in a *conditional* way, i.e. as a tendency to react in specific ways to specific internal and external conditions: 1) the motivations, moral standards, ideal values and conflicts related to these dimensions; 2) the capacities, resources and cognitive and emotional style of the subject; 3) representations of self, object and self-object relation. This information is later integrated with data on: 4) the development of these three areas and their reciprocal interaction.

The almost 800 clinicians of diverse orientation and training who took part in the SWAP-200 validation study had to use a 4-point scale to assess to what degree the items in this tool were adequate to describe the personality of their patients: 72% of the sample responded 1, the highest score, 26,7% 2, 0,6% 3 and none responded 4 (Westen & Shedler, 1999a).

The researcher or clinician who has to assess a patient with SWAP has to give a score from 0 to 7 to each of the 200 items⁷. The score of 0 is given to the items that do not describe the personality of the subject being assessed, 1 to those that describe it very little, and so on up to 7, the score to attribute to the items that are absolutely descriptive of the personality being assessed. It is important for the assessment of descriptivity not be confused with a judgement on the degree to which the trait in question is present.

The SWAP-200 procedure also requires a fixed distribution of these items, i.e. a fixed number of items to receive the various scores. This constraint, which approximates the right half of a normal curve of item distribution, is the one that comes closest to the distribution obtained by leaving the assessors free to score in the way they find most appropriate and, at the same time, is designed to avoid possible assessor bias. This term refers to the possible tendency among some assessors to give average or extreme scores independently of the real descriptivity of an item.

SWAP-200 can be carried out by someone who knows the person to be evaluated well, i.e. who has had 3-5 interviews or has had the chance to observe/listen to/read at least 3-5 video or audio recordings or transcripts of interviews. Westen and colleagues have also developed a *Clinical Diagnostic Interview* (CDI) (Westen & Muderrisoglu, 2003), similar to a group of 3 interviews of taking on charge purposely designed to facilitate SWAP.

This more clinical part of the SWAP procedure starts from the assumption that clinicians who have studied and had specific training possess more skill to listen, observe and infer patients' behaviour and psychic processes than the patients themselves or than laymen.

⁷Before the construction of the electronic support accompanying the volume *Personality Assessment with SWAP-200. La valutazione della personalità con la SWAP-200* (Westen, Shedler, & Lingardi, 2003), this scoring was done by placing the items in 7 piles.

This is why for personality assessment one cannot rely on self-report tools, which by definition cannot take into account psychic processes and implicit representations, nor the possible agreement or discrepancy between these processes and representations and those that are explicit, all of which are fundamental aspects of the assessment of personality and its disorders. To make this point clearer, suffice it to think that a self-report instrument cannot take into account the implicit omnipotence often associated with an explicit sense of personal impotence in so-called “covert”, hyper-vigilant, (Gabbard, 1994) or thin-skinned (Rosenfeld, 1987) narcissistic patients. Similarly, a narcissistic patient in terms of the DSM-IV would hardly respond “true” to an item asking if s/he has an exaggerated sense of her/his own importance. From this point of view, SWAP-200 is an attempt to operationalize and make statistically reliable and objective the procedure of assessment normally used in the clinical context. Incidentally, some of the data collected in the process of constructing the SWAP empirically supports the idea that clinicians with different training, orientation and experience are capable of reliably observing, describing and inferring the same personological configurations (Westen & Weinberger, 2004) Once the scoring of the items is finished, the SWAP-200 programme calculates and standardizes in T points (mean 50 and variance 10) the correlations between the SWAP-200 profile of the patient being assessed, that is, the set of 200 items with the scores assigned by the assessor to each one, and:

1. SWAP-200 prototypes of ideal patients with the personality disorders present in Axis II of DSM-IV, or *PD factor*,
2. prototypes of personality style empirically derived with SWAP-200 based on the descriptions of real patients, or *Q factor*.

The PD and Q factors of SWAP

The PD factors of SWAP-200 are the descriptions of *ideal patients* with a certain Axis II disorder of the DSM-IV elaborated by about 237 clinicians with different orientations and training, using SWAP-200. Besides these descriptions, the clinicians involved in this study were also asked to use SWAP to describe an ideal, “high functioning”, healthy patient.

Once the assessment of a patient is finished, the computer programme elaborates a correlation between this assessment (i.e. the 200 items and relative scores given by the rater) and the single PD prototypes (PD scores), standardizes it in T points (PD-T) and presents the results both numerically and in the form of bar-graphs. Any correlation over the score of 60 can be considered indicative of a *personality disorder*, i.e. above the mean by at least one standard deviation. If this correlation is between 55 and 60 one can talk about the presence of *strong traits* of the personality consistent with that type of disorder. In this way, SWAP-200 enables us to obtain a diagnosis that is *both categorical* (presence/absence of a disorder) *and dimensional* (what is considered is the correlation between the patient’s SWAP-200 description and *all* the prototypes of disorder of DSM-IV Axis II redescribed with SWAT-200). Besides this, there is also the calculation of the correlation between the SWAP description of the patient being assessed and the “high functioning” factor, which enables the *resources* of the personality described to be assessed. The high functioning factor, unlike the *Global Assessment of Functioning Scale* (GAF), is an index of good functioning of a strictly psychological nature, assessed by means of items like: “Is able to maintain a loving relationship characterized by authentic intimacy and by the capacity to take care of the other person” (32) or: “Can hear news that is threatening on an emotional level (i.e. information that challenges beliefs and perceptions of self and others that for her/him are fundamental) and can use it and benefit from it” (82). These differences between the PD factor of high functioning and the GAF among the aspects investigated account for the fact that the correlation between the two tools has proved positive and significant ($p = .001$), but low (.48).

These prototypes, or PD factors, showed a good convergent and discriminating validity (Westen & Shedler, 1999a) when they were correlated with the composite SWAP-200 descriptions of real patients with personality disorders classifiable in the DSM. It was also

found that the two PD factors that correlated most with the high functioning factor, i.e. the type of disorders that in our society appear more “functional”, are the obsessive and the narcissistic (*ibidem*).

The Q factors of SWAP-200 (Westen & Shedler, 1999b) were extracted with a procedure called Q-analysis from the SWAP-200 descriptions of 496 real patients who had received an Axis II diagnosis. Q-analysis is a variant of factorial analysis which, instead of joining together groups of intercorrelated items, joins together groups of cases characterized by similar scores on the same variables; in the past this procedure was used to draw up zoological taxonomies and for the study of “normal” personalities.

The application of Q-analysis to these cases has brought out the presence of 7 Q-factors. The first, within which about 20% of the subjects in the sample fell, (prevalent among them, people who in DSM had been diagnosed with depressive disorder, borderline, self-thwarting and dependent personality), underwent a second order Q-analysis and further subdivided into 5 sub-factors, for a total of 11 Q factors. The seven primary factors are:

- 1) dysphoric,
- 2) antisocial,
- 3) schizoid,
- 4) paranoid,
- 5) obsessive,
- 6) histrionic and
- 7) narcissistic⁸.

The secondary factors deriving from that of dysphoric are:

- 1a) avoidant,
- 1b) depressive-neurotic,
- 1c) dependent-masochistic and
- 1d) externalizing aggressivity, hostile.

In the case of Q factors, too, the computer programme (Westen, Shedler & Lingardi, 2003) calculates the correlation between the SWAP-200 profile of the patient being assessed and the various Q factors and standardizes it in T points (Q-T factor). When this correlation is over 60, one can talk about *personality disorder*, when it is between 55 and 60, about *strong traits* of the personality. Otherwise, one simply talks about *personality style*. It is therefore possible, also with Q factors, to obtain both a categorical and a dimensional personality diagnosis

This classification, empirically derived and clinically useful, has the limit of being elaborated on the basis of patients who could receive, and in fact did receive, a DSM diagnosis; further studies also considered patients treated in a clinical setting for “dysfunctional patterns of cognition, emotions, motivation and behaviour” who could not receive a DSM diagnosis. The results of these studies are currently being elaborated.

The classification of personality styles into Q factors showed a good convergent and discriminating validity when it was compared to clinical assessments using a 7-point scale on the extent to which the patients in the study could receive a DSM diagnosis, and based on the correlation between Q diagnosis and GAF scores. The latter data made it possible to show that the Q factors that correlate most with high functioning are the depressive, obsessive and narcissistic (Westen & Shedler, 1999b). Also the predictive validity of the tool concerning measurements like ‘number of hospitalizations after assessment’ proved to be good.

Lastly, to identify the essential dimensions that must be taken into account to draw up an adequate personality profile, a factorial analysis was carried out on the SWAP-200 items. The results of this analysis showed the following orthogonal factors:

- 1) psychological health,
- 2) psychopathy,

⁸Later research into SWAP descriptions of 122 adult patients showed the presence of 3 narcissistic sub-types: grandiose/malign, fragile and high functioning (Russ, Bradley, & Westen, in preparation).

- 3) hostility,
- 4) narcissism,
- 5) emotional dysregulation,
- 6) dysphoria,
- 7) schizoid orientation,
- 8) obsessiveness,
- 9) thought disorder /schizotype,
- 10) Oedipal conflicts /histrionic sexualization,
- 11) dissociation,
- 12) sexual conflicts.

The identification of these factors, which cannot be further reduced, challenges the clinical adequacy of tools that in other respects are sophisticated and reliable (like the various questionnaires connected to the Five Factor Model), but which do not seem able to account for an adequate number of psychopathologically relevant dimensions (Shedler & Westen, 2004).

SWAP-200 and SWAP-II have show an average level of interrater reliability and test-retest reliability of .80 (Westen & Muderrisoglu, 2003) and the SWAP diagnoses have proved to be predictive of dimensions like adaptive functioning, suicide history, hospitalizations, arrests, involvement in abusive relationships, development history, family history and patient response to treatments. It therefore seems possible, to use Westen's expression, to use tools like SWAP to transform empirical data into clinical gold. In other words, tools like SWAP seem to enable us to perfect current diagnoses, empirically derive new diagnoses, empirically derive personality traits using a different method from the self-report, to identify diagnostic sub-groups of patients who share the same diagnostic label (eg, abusive husbands, narcissistic sub-types), verify aetiological hypotheses, construct empirically based scales for specific, clinically relevant purposes (eg, to predict responses to treatment, legal risk, etc) and to measure the change in character in the course of psychotherapies (Lingiardi, Shedler, & Gazzillo, 2006). Let us see some examples of this research.

Clinically relevant empirical research

SWAP procedure, briefly described in the previous section, has proved adequate for the needs of empirically sound research and at the same time clinically relevant. Some research has in fact demonstrated the possibility of predicting a *composite risk factor* (arrests, violent crimes, beating) related to the behaviour of adults of the male sex on the basis of the SWAP-II diagnosis of their personality traits. In particular, the personality trait *psychopathy* correlates in a positive, significant way with this factor ($r = .62$ with $p < .001$), while the *narcissistic* trait seems to be a factor of resilience towards these risks ($r = -.12$ with $p < .004$). Other relevant predictors of arrests and criminal or violent behaviour are *childhood psychopathy* ($r = .14$ with $p < .001$) and the presence of *multiple traumas at an early age* ($r = .15$ with $p = .001$).

Other research conducted with SWAP on samples of adolescents and adults (Thompson-Brenner et al., in preparation; Thompson-Brenner & Westen, 2005; Westen & Harnden-Fischer, 2001) have allowed three personality sub-types to be identified in patients with *eating disorders*. Taking these three sub-types into account – one of high functioning, one constricted and one dysregulated – replicated with different tools, informers and statistical procedures, it is possible to predict the duration and the outcome of psychotherapy with these patients.

In particular, patients with eating disorders and dysregulated personality (i.e. with a borderline personality and emotional dysregulation, but ego-dystonic) need longer therapies to obtain the disappearance of the eating disorder ($r = .41$ with $p < .001$), while the overall outcome of psychotherapy is often negative both for patients with eating disorder and dysregulated personality and for those with eating disorder and constricted

personality (respectively $r = -.24$ with $p < .05$ and $r = -.19$ with $p < .05$), but not for those with eating disorder and high functioning personality.

In cases of adolescents with symptoms of bulimia, then, the presence of a high functioning personality allows for a positive outcome to the psychotherapy ($R = .61$ with $p < .001$), while where there is a dysregulated personality an unfavorable diagnosis seems to be implied ($r = -.31$ with $p < .05$). Incidentally, this data seems to seriously challenge the Est studies which, as we have mentioned, try to assess the efficacy of psychotherapy for Axis I disorders without considering the conditions related to Axis II.

Similar research was conducted with adult patients who had been diagnosed with a generalized anxiety disorder (Peart & Westen, in preparation). Using the SWAP descriptions of real patients with GAD provided by 201 clinicians (American psychologists and psychiatrists of different orientations) who had them in treatment, 4 personality sub-types were identified: dysphoric, high functioning, emotionally dysregulated, obsessive. Based on these sub-types it was possible to draw up predictions about the presence of a comorbidity or substance abuse, the global level of functioning, the presence of breakdowns in the attachment system in childhood and a history of early traumas, as well as the outcome of the psychotherapy.

In particular, a lower global functioning is typical of patients with GAD and dysregulated personality ($r = -.42$ with $p > .001$), who are also the ones that most often abuse substances ($r = .30$ with $p < .001$), suffered traumas and attachment system breakdowns in childhood ($r = .20$ and $r = .22$ with $p < .05$) and present an unfavorable prognosis in psychotherapy ($r = -.24$). For all these indicators, patients with GAD and high functioning have a good prognosis in psychotherapy ($r = .37$ with $p < .001$) and in general did not suffer traumas or childhood attachment breakdowns ($r = -.32$ and $r = -.28$ with $p < .001$).

Other research worth noting was carried out with teenage patients in therapy with the "questionable" diagnosis of attention deficit disorder and hyperactivity (Levin & Westen, in preparation), a syndrome at the centre of debates involving professionals who believe in its existence and hope for an early pharmacological solution against professionals who question the very meaning of the diagnostic label, which seems to describe a cluster of behaviours lacking sense unless they are seen in the personal and social context in which they are manifested. Q factor analysis applied to SWAP descriptions of 137 teenagers with ADHD carried out by the clinicians who had been treating them in psychotherapy showed the presence of 4 personality sub-types: psychopathic, socially withdrawn, emotionally dysregulated and high functioning. The patients with ADHD and emotionally dysregulated personality present adverse events, traumas and attachment breakdowns in childhood ($r = .25$, $r = .21$, $r = .24$ with $p \leq .05$); in their family history there are often diagnoses of psychotic disorders, mood disorders, symptoms of exteriorization or suicide in close relatives ($r = .20$, $r = .17$, $r = .24$ e $r = .21$ with $p \leq .05$); all these conditions are generally absent in adolescents with ADHD and high functioning personality.

The level of global functioning tends to be high in patients with ADHD and high functioning personality ($r = .41$ with $p < .001$) and low in those with emotionally dysregulated personality ($r = -.35$ with $p < .001$). A year after the beginning of the psychotherapy, the outcome on ADHD symptoms tends to be negative in socially withdrawn patients ($r = -.29$ with $p < .02$) and those with psychopathic personality ($r = -.27$ with $p = .05$), but not in those with high functioning personality.

It should be stressed that it is impossible to make any of these distinctions if the patients with ADHD are considered as a single group; in other words, it is essential to know the personality that constitutes the diathesis of the Axis I disorders in order to obtain all the information just summarized. The EST research, however, systematically fails to investigate Axis II comorbidity in the target disorders.

Remarks on research into transference and countertransference

A further example of Westen's approach to research on issues of clinical relevance is that of the empirical study of transference and countertransference. 181 American clinicians of

different training and orientation used the *Psychotherapy Relationship Questionnaire* (PRQ) (Bradley, Heim, & Westen, 2005) and the *Countertransference Questionnaire* (CTQ) (Betan, Heim, Conklin, & Westen, 2005) to describe their subjective experience of the relationship with one of their patients chosen at random (the last patient seen the week before receiving the research material). Although it is difficult to maintain that questionnaires compiled by clinicians, rating scales and factorial analyses can provide information about transference and countertransference (which by definition are unconscious processes and largely idiomatic), some of the results of this research seem quite interesting.

Firstly, the factorial analysis of data on the PRQ showed that the relationship of the patient with the therapist – as it is perceived by the clinician – can be described using 5 broad dimensions:

- 1) secure/engaged,
- 2) avoidant/counterdependent,
- 3) anxious/preoccupied,
- 4) angry/entitled,
- 5) sexualized.

The therapist's possible emotional attitudes towards patients (CTQ) can on the other hand be described on 8 broad dimensions:

- 1) overwhelmed/disorganized,
- 2) helpless/inadequate,
- 3) positive,
- 4) special/overinvolved,
- 5) sexualized,
- 6) disengaged,
- 7) parental/protective,
- 8) criticised/mistreated.

At least 3 of the dimensions highlighted by the PRQ and 4 of those underlined by the CTQ recall the attachment styles hypothesised in adults on the basis of data collected with the *Adult Attachment Interview* (AAI) (Main et al., 2002), which encourages researchers and clinicians who believe the construct of *attachment style* to be a useful contribution also for the description of clinical situations.

Conclusions

Apart from some conceptual simplifications, inevitable when one tries to translate a clinical construct into an empirically assessable form, Westen's contribution to clinical research seems extremely important.

The idea of a "return to the future" (Westen, Gabbard, & Baglov, 2006) which: a) recovers the approach to the study of psychopathology for which personality and its structure constitute the diathesis for possible Axis I problems; b) does not neglect the implicit dimensions and unconscious psychic processes that contribute to personological functioning; and c) tries to join classification with description and understanding of psychic processes, is already bearing fruit.

From the methodological point of view, using a working network that can involve professionals with different origins, training, orientation and experience and the attempt to give clinicians the task of doing what they know best (observing and making inferences) and researchers that of elaborating and analysing the data in a methodologically sound manner, are strategies that have already proved to bring important results.

SWAP-200 – only to mention Westen's most well-known and successful tool – has been translated and is in use in many countries (USA, Italy, Spain, Mexico etc.), and it is among the preferred tools for the assessment of psychic functioning recommended by the *Psychodynamic Diagnostic Manual* (PDM, 2006). SCORS (Westen, 1991) made it possible to collect relevant information in assessing the differential impact of physical and sexual abuse on children and for the understanding of the subtle differences between

depressed, borderline and normal subjects through the aspects of representation of self and others and of social cognitions.

To conclude, Westen (2002), by not using the authority principle to legitimize statements, but also by distancing himself from certain postmodernist epistemological trends that say, "I know nothing, but you don't know anything either", sums up his approach in the following way: "The antidote to authorized knowledge is a scientific attitude capable of recognising that scientific knowledge is always imperfect, probabilistic and influenced by extra-scientific forces – social, economic and psychodynamic – but also that this imperfect knowledge, understood in its human context, is better than knowledge sanctioned by orders or legitimated by nihilism" (p. 892).

References

American Psychiatric Association (1987). *Diagnostic and Statistic Manual of Mental Disorders* (3rd ed.). Washington, D.C.: APA.

American Psychiatric Association. (1994). *Diagnostic and Statistic Manual of Mental Disorders* (4th ed.). Washington, D.C.: APA

Betan, E., Heim, A. K., Conklin, C. Z., & Westen, D. (2005). Countertransference phenomena and personality pathology in clinical practice: An empirical investigation. *American Journal of Psychiatry*, *162*, 890-898.

Bradley, R., Heim, A., & Westen, D. (2005). Transference phenomena in the psychotherapy of personality disorders: An empirical investigation. *British Journal of Psychiatry*, *186*, 342-349.

Chambless, D., & Ollendick, T. (2000). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, *52*, 685-716.

Eysenck, H. J. (1952). The effects of psychotherapy. *Journal of Consulting Psychology*, *16*, 319-324.

Gabbard, G. (1994). *Psychodynamic Psychiatry*. American Psychiatric Press.

Goldfried, M. (2000). Consensus in psychotherapy research and practice: Where have all the findings gone? *Psychotherapy Research*, *10*, 1-16.

Grünbaum, A. (1984), *The foundation of psychoanalysis*.

Howard, K. I., Cornille, T., Lyons, J. S., Vessey, J., Lueger, R. J., & Saunders, S. M. (1996). Patterns of mental health service utilization. *Archives of General Psychiatry*, *53*, 696-703.

Howard, K., Kopta, S., Krause, M., & Orlinsky, D. (1986). The dose-effect relationship in psychotherapy. *American Psychologist*, *41*, 159-164.

Kazdin, A. (2002). Overview of research design issues in clinical psychology. In P.C. Kendall, J. Butcher, & G. Holmbeck, (Eds.), *Handbook of Research Methods in Clinical Psychology* (pp. 330-363). New York: John Wiley & Sons.

Kendall, P. C. (1998). Empirically supported psychological therapies. *Journal of Consulting and Clinical Psychology*, *66*, 3-6.

Kendall, P. C., Marrs-Garcia, A., Nath, S. R., & Sheldrick, R. C. (1999). Normative comparisons for the evaluation of clinical significance. *Journal of Consulting and Clinical Psychology*, *67* (3), 285-299.

Kopta, S., Howard, K., Lowry, J., & Beutler, L. (1994). Patterns of symptomatic recovery in psychotherapy. *Journal of Consulting and Clinical Psychology*, *62*, 1009-1016.

- Levin, D., & Westen, D. (in press).
- Lingiardi, V., Shedler, J., & Gazzillo, F. (2006). Assessing Personality Change in Psychotherapy with the SWAP-200: A Case Study. *Journal of Personality Assessment*, 86, 23-32.
- Main, M. et al. (2002), *Adult Attachment Interview Coding Manual*. Unpublished manuscript.
- Messer, S. (2001). Empirically supported treatments: What's a non-behaviorist to do? In B. D. Slife, R. N. Williams, & S. H. Barlow (Eds), *Critical issues in psychotherapy: Translating new ideas into practice* (pp.21-25). Thousand Oaks, CA: Sage.
- Nakash-Eisikovits, O., Dutra, L., & Westen, D. (2002). The relationship between attachment patterns and personality pathology in adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41 (9), 1111-1123.
- Nathan, P. E., Stuart, S. P., & Dolan, S. L. (2000). Research on psychotherapy efficacy and effectiveness: Between Scylla and Charybdis? *Psychological Bulletin*, 126 (6), 964-981.
- Newman, D. L., Moffitt, T., Caspi, A., & Silva, P. A. (1998). Comorbid mental disorders: Implications for treatment and sample selection. *Journal of Abnormal Psychology*, 107, 305-311.
- Peart, J.L., & Westen, D. (in preparazione).
- PDM Task Force (2006). *Psychodynamic Diagnostic Manual*. Silver Spring, MD: Alliance of Psychoanalytic Organizations.
- Rosenfeld, H. A. (1987). *Impasse and interpretation*. London: Tavistock
- Russ, E. Bradley, R., & Westen, D. (in preparazione)
- Seligman, M. (1995). The effectiveness of psychotherapy. *American Psychologist*, 50, 965-974.
- Shedler, J., & Westen, D. (2004). Dimensions of personality pathology: An alternative to the Five Factor Model. *American Journal of Psychiatry*, 161, 1743-1754.
- Thompson-Brenner, H., & Westen, D. (2005). A naturalistic study of psychotherapy for bulimia nervosa, Part 1: Comorbidity and treatment outcome. *Journal of Nervous and Mental Disease*. 193, 573-584.
- Thompson-Brenner, H., et al. (in preparazione)
- Westen, D. (1990). Toward a revised theory of borderline object relations: Implications of empirical research. *International Journal of Psycho-Analysis*, 71, 661-693.
- Westen, D. (1998). Case formulation and personality diagnosis: two processes or one? In Barron, J. (Ed.), *Making Diagnosis Meaningful: Enhancing Evaluation and Treatment of Psychological Disorders* (pp. 111-137). Washington, DC: American Psychological Association.
- Westen, D. (2002). The language of psychoanalytic discourse. *Psychoanalytic Dialogues*, 12 (6), 857-898.
- Westen, D., Gabbard, G., & Blagov, P. (2006). Back to the future: Personality structure as a context for psychopathology. In R.F. Krueger & J. L.Tackett (Eds.), *Personality and psychopathology* (pp.335-384). New York: Guilford.
- Westen, D., & Harnden-Fischer, J. (2001). Personality profiles in eating disorders: Rethinking the distinction between Axis I and Axis II. *American Journal of Psychiatry*, 158, 547-562.
- Westen, D., & Morrison, K. (2001). A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 69 (6), 875-899.

- Westen, D., & Muderrisoglu, S. (2003). Reliability and validity of personality disorder assessment using a systematic clinical interview: Evaluating an alternative to structured interviews. *Journal of Personality Disorders, 17*, 350-368.
- Westen, D., & Shedler, J. (1999a). Revising and Assessing Axis II, Part I: Developing a clinically and empirically valid assessment method. *American Journal of Psychiatry, 156*, 258-272.
- Westen, D., & Shedler, J. (1999b). Revising and assessing Axis II, Part II: Toward an empirically based and clinically useful classification of personality diagnosis. *American Journal of Psychiatry, 156*, 273-285.
- Westen, D., Shedler, J., & Bradley, R. (in press), A Practice Research Network Approach to the Classification of Personality Pathology in Adults and Adolescents: "Virtual Field Trials".
- Westen, D., Shedler, J., & Lingardi, V. (2003). *La valutazione della personalità con la SWAP-200*. Milano: Raffaello Cortina.
- Westen, D., & Weinberger, J. (2004). When clinical description becomes statistical prediction. *American Psychologist, 59*, 595-613.
- Wilkinson-Ryan, T., & Westen, D. (2000). Identity disturbance in borderline personality disorder: An empirical investigation. *American Journal of Psychiatry, 157*, 528-541.
- Wilson, G. (1998). Manual-based treatment and clinical practice. *Clinical Psychology: Science & Practice, 5* (3), 363-375.
- Zimmerman, M., McDermt, W., & Mattia, J. (2000). Frequency of anxiety disorders in psychiatric outpatients with major depressive disorder. *American Journal of Psychiatry, 157*, 1337-1340.
- Zittel, C., Bradley, R., & Westen, D. (2006). Affect regulation in borderline personality disorder. *Journal of Nervous and Mental Disease, 194*, 69-77.