Fonagy, P; (2010) The changing shape of clinical practice: Driven by science or by pragmatics? **Psychoanalytic Psychotherapy** , 24 (1) 22 - 43. 10.1080/02668731003590139.

RESEARCH ARTICLE

The changing shape of clinical practice: driven by science or by pragmatics?

Peter Fonagy, PhD, FBA

Freud Memorial Professor of Psychoanalysis, University College London

Chief Executive, The Anna Freud Centre, London

Research Department of Clinical, Educational and Health Psychology, University College London, Gower Street, London WC1E 6BT, p.fonagy@ucl.ac.uk

Paper submitted to *Psychoanalytic Psychotherapy*, December 2009.

Introduction

Few would dispute that psychoanalytic practice has changed considerably over the last 50 years. The differences are at least threefold: (a) a bimodal distribution of treatment lengths (some very long and some very short); (b) patients with personality pathology in addition to axis I symptoms; and (c) technical changes with a relational flavour, displaying influences from developmental theories including attachment theory. Are any of these changes associated with scientific research, and if so, how? A second question concerns the research base of psychoanalysis. If we take the discoveries of the last twenty five years seriously, what kind of therapy might we recommend to our patients? Would it look anything like standard practice? I will attempt to pose some serious questions about the status of clinical practice, asking why it appears to be protected from scientific advances yet readily responsive to pragmatic considerations such as third party payments. To make the argument easier to absorb I will try and tell it as number of short heroic tales, each with its own beginning, middle and end complete with the moral lessons to be learned.

The evidence base of psychotherapy in general

I will begin with the story of the evidence base of psychotherapy in general. This story began over half a century ago when an evil genius, a certain Dr Hans Eysenck (1952) disturbed the tranquility in the Kingdom of Psychotherapy established by a benign powerful ruler (Sigmund Freud, the Wise). This evil genius made the claim that psychotherapy worked no better than the facility

nature gave all of us to recover from psychological disturbance by a process of spontaneous healing. Up to that time, all the Templar Knights of psychotherapy, who were both fearsome warriors and devout monks, could righteously believe that the people they cured recovered because of their magic spells and carefully measured potions, which took often decades of apprenticeship to learn to brew with confidence. This evil genius claimed that Sleeping Beauty woke up simply because she had slept long enough, and that her awakening had nothing to do with the magical time she spent on the couch.

Fortunately, in raising this monster, he disturbed a nest of psychologists, who saw an opportunity to create a new hive for their worker bees (also called graduate students) and they called this psychotherapy research. The researcher bees and their bee keepers went on to sample hundreds of thousands of therapy sessions and collected together a number of papers with magical powers, called an 'evidence base' which showed that the magic of psychotherapy was indeed magical.

In the ensuing half a century the bee keepers and the researcher bees (the bee keepers actually looked more like Sancho Panza, and the templar knights a little like Don Quixote), demonstrated that the average EFFECT SIZE of psychotherapy was .8 across probably over 1,000 studies (Wampold, 2001). Effect size (ES) is researcher bee language for the likelihood that a person treated with psychotherapy magic would be better off than a person in the control

group if both were chosen at random (Cohen, 1962). Using ESs the bees can collect together the nectar (the results) of hundreds of flowers (studies) and calculate the average ES for a treatment. Tradition has it that an effect size of less than 0.3 is too close to the 50% mark (i.e. too close to chance), to call it a significant effect. From 0.3 (where 58% of those treated are better off than untreated controls) to 0.6 (where 2/3 are better off) the effect is considered weak. From 0.6 to 0.9 (where ¾ of those treated are better than untreated controls) effect sizes are thought of as medium and beyond 0.9 we are in the land of strong effect sizes. You should note that effect sizes are sometimes used to describe the change between the beginning and the end of a trial (pre-post effect size) and the difference between a treated and an untreated group at termination (between group effect size). Pre-post effect sizes combine the effect of treatment with the monster that Eysenck raised – spontaneous remission. One final note about the language of these researcher bees – in another of the dialects spoken in this evidence based magical beehive, differences are expressed in terms of the number of individuals who need to be treated before we come upon a person who unequivocally would not have got better on their own without the treatment (Laupacis, Sackett, & Roberts, 1988).

So, 0.8 as a psychotherapy versus no therapy effect size is BIG -- in a moderate sort of way (Wampold, Imel, & Minami, 2007). It means that nearly ¾ of patients who have psychotherapy are better off than those left to recover by themselves. The Number Needed to Treat (NNT) is 3. As a marker, for those of you taking

aspirin as a prophylaxis for heart attacks, the NNT before you see one unequivocal case who benefitted is 129. The effect size for Psychotherapy Magic is superior to almost all interventions in cardiology, geriatric medicine, asthma, flu vaccines, cataract surgery. Psychotherapy is mostly as effective as psychoactive medication and there is evidence that additional benefit accrues from combining the magics in many contexts (e.g. Cuijpers, van Straten, van Oppen, & Andersson, 2008; Maina, Rosso, & Bogetto, 2009).

[Page 24 →]

So this story appears to have a happy ending. But like most of the stories that you will hear from me, and unlike real fairy stories, the ending is also curious and complex. Effect sizes are inherently ambiguous. A magical treatment could attain a moderate effect size by producing a very LARGE EFFECT for a SMALL SUBSET of patients or achieving a moderate but incomplete reduction in symptoms for many. To clarify this, the researcher bees have collected data on the percent recovered or percent improved both by some predefined criterion. Two of the busiest beekeepers of psychotherapy research, Drew Westen and Elizabeth Bradley (Westen & Bradley, 2005), have provided us with an impressive meta-analysis of six disorders (MDD, panic disorder, GAD, OCD, bulimia nervosa and PTSD) based on percent improvers. They tell a sobering tale. Whilst across disorders, treatment versus control effect sizes tend to be moderate or large, on average only roughly half of patients who *complete*

treatments in these trials improve substantially. The figures are even more sobering when we compute the percent of those who improved on the basis of all those who *entered* treatment. On average, only 33% of those who entered treatment for depression can expect to improve significantly based on these figures. A further surprise awaits us when we look at the percent who remain improved at follow-up. Less than a third of those that complete treatment for bulimia remain improved. But the figures for anxiety conditions (panic and GAD) tend to be better.

Not surprisingly, improvement rates relate to severity and treatment duration (Kopta, Lueger, Saunders, & Howard, 1999). On average, acute distress improves in ¾ of cases within 25 sessions. But chronic disorders, defined in various ways, appear to require longer term treatment. There is a less than 60% improvement rate after 25 sessions. The situation is even worse for those who receive what we might refer to as a complex disorder diagnosis, that is, 3 or more diagnoses or diagnosis of a personality disorder. Here improvement rates are less than 40% after 25 sessions. There is even some indication, good magic turning bad, of inadvertent harm being done to patients with some personality disorders if they are offered time-limited treatment, so that they end up worse off than when they started (Tyrer et al., 2004). The most recent German work from the bees in Horst Kaechele's hive, a champion beekeeper of psychotherapy research, carefully following a group of patients with private health care insurance, reported that about one and a half years of treatment was required

before the average patient achieved an acceptable level of improvement (Puschner, Kraft, Kachele, & Kordy, 2007). Incidentally, this work found no evidence for the much cited exponential rate of improvement originally demonstrated by Ken Howard (Howard, Kopta, Krause, & Orlinsky, 1986). The relationship of improvement and time is best illustrated by a straight line.

So what are the morals of the tale of psychotherapy in general? Certainly there is good evidence for the efficacy of psychological therapies. Over 1,000

[Page 25 →]

studies demonstrate that in relation to major mental health conditions they can achieve significant symptom reductions and in some cases, particularly with anxiety disorders, freedom from symptoms. The magic spell and potion, when appropriately administered, can improve social adjustment and work relationships. However there are many different potions and some commonly used ones have been less well researched than others. There is more evidence for the magic formula brewed by the knights of cognitive behaviour therapy than the potion generated in the cauldron of psychodynamic psychotherapy. But of course, in practice most of those who brew these powerful mixtures combine several different recipes, and currently there is very little evidence for eclectic brews like these. By contrast, researcher bees have been busy in relation to some little used concoctions, such as interpersonal psychotherapy. There are

also many diagnoses for which very little data has been collected, including eating disorders, bipolar disorder, and certain personality disorders such as avoidant PD.

The story of long-term psychodynamic psychotherapy

The next story is less epic and more recent than the previous one. As with many modern stories, it is fraught with controversies and possibly insoluble problems. The beekeepers and researcher bees struggle to collect information about longer-term treatments to bring back to our collective knowledge base. They found it hard to solve the problem of randomization when this involved getting agreement to go without the preferred treatment for 18 months or more. Finding an appropriate control group was itself a challenge, as was the blinding of assessors, and the creating of manuals to guide work over a lengthy time period and ensure that what took place was what was described (treatment integrity). Yet for psychoanalytic clinicians this is the Holy Grail. It is long term treatment that we have been trained to do and that we wish to make claims for.

It is nothing short of miraculous that enough data have been collected for two prestigious systematic reviews to have been published, one by our friend Falk in the Journal of the American Medical Association,(Leichsenring & Rabung, 2008) and another by Askia de Maat and her colleagues in the Harvard Review of Psychiatry (de Maat, de Jonghe, Schoevers, & Dekker, 2009). The latter review collected together 27 studies where the impact of long term therapy on symptom

reduction was measured, and/or information on personality changes was collected. The unmet challenge of the control groups meant that effect sizes were pre-post, not between-group. Nevertheless, the studies covered the treatment of over 5,000 patients. The effect sizes of outcome measures combined were between 0.8 and 1 and tended, if anything, to slightly increase on follow-up and were somewhat bigger for psychoanalysis than psychotherapy. The percent success rate on symptoms was around 70% based on clinicians' opinion and between 60-70% for patient self-report, when success was defined as at least moderate improvement.

[Page 26 →]

The Leichsenring meta-analysis was very ambitious and identified 23 studies. The studies concerned difficult problems, but pre-post effect sizes were consistently large. For chronic problems, the effect size is 0.87 - 2.45, for complex depression and anxiety it is 0.97-1.94, multiple problems 0.94-1.84, and personality disorder 0.82-1.65. Controversially, the authors contrasted these effects with those normally obtained for similar client groups in short-term therapy and found a significant superiority for long term treatment. The size of effects varied according to type of measure, with the largest effect sizes consistently obtained on target problems, and social and personality functioning coming some way behind. However the effects were consistently positive, with the confidence

interval around the effect sizes comfortably above the line of insignificance, the dreaded 0 line.

As you might imagine, not all in the Land of Psychotherapy Research were happy on hearing this news (Beck & Bhar, 2009; Glass, 2008; Kriston, Holzel, & Harter, 2009; Roepke & Renneberg, 2009; Thombs, Bassel, & Jewett, 2009) and the ensuing correspondence had the consistent theme of debunking positive evidence for long term treatment. The details should perhaps be consigned to the history books, but the points raised address both the nature of the original studies (the review was based on studies with small samples with a likely bias towards the positive, treating a wide range of disorders with poorly specified control groups, and uncontrolled for contact and the structure of treatment) and the methodology of the review (conflating within-group and between-group effect sizes, selective inclusion and exclusion of studies etc).

Our bards put up a spirited defense over several pages of the journal (Leichsenring & Rabung, 2009). But the substance of the attacks was hard to deny. Many of the studies reviewed were in effect uncontrolled and heterogeneous and it is hard to feel confident that the original critique by the Evil Genius, Hans Eysenck, was adequately addressed by the original collection of reports. So the bards returned to their crypts and wrote a new ballad which is not yet published, and therefore I am only at liberty to share the refrain. They identified 10 controlled studies of long term psychoanalytic psychotherapy versus

other types of treatment (Bachar, Latzer, Kreitler, & Berry, 1999; Bateman & Fonagy, 1999; Bateman & Fonagy, in press; Clarkin, Levy, Lenzenweger, & Kernberg, 2007; Dare, Eisler, Russell, Treasure, & Dodge, 2001; Gregory et al., 2008; Huber, Denscherz, Gastner, Henrich, & Klug, submitted; Korner, Gerull, Meares, & Stevenson, 2006; Svartberg, Stiles, & Seltzer, 2004) where these treatments were used in the treatment of complex disorders, chiefly personality disorder (7), eating disorders (2) and depression (1). The comparisons are with CBT, DBT, CAT, SCM and TAU. The treatments lasted on average 70 weeks offering 120 sessions. The comparison treatments lasted about the same time, although offered fewer sessions. The findings were similar to the previous analysis. The average between-group effect

[Page 27 →]

size was 0.67, somewhat larger for target problems, 0.88, than general psychiatric symptoms, 0.54. Medium effect size differentiated the comparison groups from LTPP in terms of personality and social function.

Why are these findings of enormous importance? This is the first set of strong signals which suggests that long term psychodynamic psychotherapy is superior to less intensive treatments when directed towards complex mental disorders. No doubt when these findings are published there will be a chorus of complaints concerning the original studies and the methodology of the review. But should

this surprise us? We live in a competitive world. What is good for psychodynamic therapy is often felt to disadvantage other orientations. Collectively, we have expressed dissatisfaction on many occasions about similar issues in relation to trials of CBT.

Nevertheless, before concluding, let us once again draw the morals from our tale. First, whilst it is reassuring and helpful that LTPP knights struggle for longer and ultimately more effectively than the knights with other crests on their shields, we do not know how well the other knights would do if their rules of combat permitted them to battle as intensely and at as close a range as our knights did. Second, the knights indeed showed their valour in these most ardent of trials rescuing suicidal self-harming damsels in distress (both BPD and eating disorder samples are 80% female). But how common are such challenges in our consulting rooms where we take the same amount of time, and sometimes longer, to rescue individuals of both genders, who perhaps at least superficially are in less distress? Is the superiority of our knights still evident, and if so in what respect? The study of private insurance cases I mentioned earlier (Puschner et al., 2007) would suggest caution. When psychoanalytic psychotherapy and psychodynamic psychotherapy were tracked over two years in 480 patients no significant differences were seen in the rate or extent of decline between these two groups. So, what is the moral? We need more trials with a broader set of outcome measures.

Before closing my tale about LTPP, let me make three further observations; first on the naturalistic follow-up of patients treated by members of this organisation in long term therapy or psychoanalysis (Beutel, Rasting, Stuhr, Ruger, & Leuzinger-Bohleber, 2004; Leuzinger-Bohleber, Stuhr, Ruger, & Beutel, 2003). It was an interesting and carefully constructed study, not so much for the findings but because the Templar knights who offered the treatment worked alongside the researcher bees and used qualitative as well as quantitative methods to show how the patients experienced their treatment. The vast majority of patients valued their experiences and the impact of treatment could be measured in terms of healthcare costs. The most ambitious study of psychoanalytic psychotherapy comes from Helsinki (Knekt et al., 2008). It contrasted Solution Focused Therapy and psychodynamic psychotherapy with long term psychodynamic psychotherapy for patients with mixed depression and anxiety problems. These beekeepers and research bees worked extremely hard and pursued their patients relentlessly over three years. This was just as well

[Page 28 →]

because significant benefit from long term treatment was not found at 18 months, nor even at 24 months but only at 36 months. In fact our knights, who had been trained to do battle for years rather than months, took their time to achieve success and were frankly struggling to keep up with the progress of their short term colleagues over the first year of the trial. But of course it could be said that

the patients who selected themselves for long term treatment were the tougher customers. The third study is probably the best study of LTPP so far carried out. The Munich study of Dorothea Huber and Gunter Klug (Huber et al., submitted) is remarkable for being the only one to use a specifically psychoanalytically oriented outcome measure, Wallerstein's Scales of Psychologoical Functioning (Klug & Huber, 2009). The instrument is based on psychoanalytic experts' definitions and includes 17 dimensions, each divided into 2 sub-dimensions, (a) exaggerated and (b) inhibited functioning. So impulse regulation may be pathological because of over-indulgence or over-inhibition. You could finally see the benefits of psychoanalysis on this measure, but even here, convincingly only after a one year-wait.

So, it seems to me that in looking at intensive long term treatment we are probably measuring the wrong things and not for long enough.

The current story of psychotherapy for BPD: an example

Before ending this part of my story and whilst I still have your attention, I would like to tell you one of my own tales. Indulge me please, not only because it is a story where I played the part of a beekeeper but because it illustrates a number of the issues of evidence with which we as knights of psychodynamic psychotherapy are having to confront.

Now, there has been many a trial with patients known as borderline and many a potion has been tried and shown to be effective in randomized controlled trials. These include DBT (e.g. Linehan et al., 2006), TFP (e.g. Clarkin et al., 2007), SFT (e.g. Giesen-Bloo et al., 2006), CBT (e.g. Cottraux et al., 2009), CAT (e.g. Chanen et al., 2008), DDP (e.g.Gregory et al., 2008), STP (e.g. Blum et al., 2008) and MBT (e.g.Bateman & Fonagy, 2008). It should be clear to you that as long as the acronym for the magic formula has three letters in it, it will be shown to be superior, mostly against treatment as usual or a less than adequate comparison. All these treatments are specialist interventions requiring extensive training and continuous supervision.

So, when Antony Bateman and I started out on our quest (Bateman & Fonagy, in press), we made solemn oaths that we were not going to return unless we found a comparison that was more meaningful to practitioners and third party payers than the comparisons we found in the journals kept by the High Priests of evidence based practice. Our solemn oath committed us to find superiority

[Page 29 →]

compared to a structured treatment organized in a coherent treatment programme with equivalent supervision, when both treatments were delivered by knights trained to the same level but without family crests committing them to one or other side. We also swore to collect a clinically representative sample of

damsels and gentlemen with confirmed diagnosis of BPD and at high risk of suicide. We swore not to stop until we reached adequate statistical power to detect even relatively small differences.

Our quest concerned the trial of Mentalisation Based Treatment (Bateman & Fonagy, 2006), a potion we ourselves concocted containing special ingredients to help patients understand their mental states more clearly, both simple and complex, and assist them in regulating their emotions, focus their attention (effortful control) and conceptualise their own actions and the actions of others in terms of thoughts, feelings, wishes, beliefs and desires. MBT is not concocted to make people 'see in the dark' (in other words it is not aimed at unconscious insight) but rather to help them to make better psychological sense of preconscious aspects of their interpersonal (especially their attachment) relationships, including their relationship with their therapist.

The trial was open to women or men who made a suicide attempt or a life-threatening act of self harm within the last six months and who carried a confirmed label of BPD. Our commitment was to treat all comers to ensure that our potion worked for all, except those with organic brain disorder, psychotic disorder or opiate dependence. We assessed patients at admission, six months, 12 months and 18 months. The programme lasted 18 months and included an individual and a group therapy session once per week.

Now, in order to ensure that it was our potion that was responsible for the healing and not our special knights with particular family crests, we co-opted 11 apprentice knights who were randomly assigned to be trained to deliver outpatient MBT or outpatient structured clinical management (SCM). Structured clinical management is what a good psychiatric department should offer these patients in a coherent and ordered manner. All therapists had a minimum of 2 years experience of treating patients in general psychiatric services following completion of their general psychiatric training and at least one year's experience of treating patients with BPD. Their average psychiatric experience was well over six years. Amongst our knights were 7 nurses (4 MBT and 3 SCM), three trainee psychiatrists (2 MBT and 1 SCM) and one accredited counselor (SCM). We trained both the MBT and SCM therapists, and both trainings were carefully designed to be of roughly equal duration, 3 days basic training and 2 days advanced training for MBT and comparable basic training on personality disorders and generic supportive psychotherapeutic techniques for SCM. Supervision was offered on a weekly basis for the knights in both arms of the trial.

SCM included many of the ingredients of MBT, including support and structure, challenging of self destructive acts, and crisis management, but in addition it included advocacy, social support work, problem solving and regular

[Page 30 →]

medication reviews. Knights trained in the MBT methods were taught how to enhance basic mentalising skills, how to offer mentalising interpretations, how to mentalize the transference, how to deal with crises, emotional storms, self-harming behavior and suicide attempts by addressing these experiences as failures of mentalising triggered by a variety of interpersonal experiences. The medication review for this group followed the same protocol as the SCM group, but the therapist knights were required to attempt to explore the psychological reasons behind the requests by the patients or others involved with the patients for changes in medication, in the same way as they would for all the patient's other actions.

168 patients were screened for eligibility and 134 were randomized. Of the patients excluded, 1/3 did not attend interviews and a further 1/3 declined participation. The rest either did not meet inclusion or exclusion criteria or were uncontactable. 52 of 71 participants allocated to MBT completed treatment and 47 out of 63 allocated to SCM did so. There was minimal missing data and all allocated patients were included in the analysis. The primary outcome was the proportion of each group without severe para-suicidal behavior as indicated by suicide attempts, life-threatening self-harm or hospital admission in the previous 6 months. The randomization was successful and the groups did not differ in key characteristics. This was a fairly complex group of patients with about 3 Axis I diagnoses and 2.5 Axis II diagnoses in addition to BPD. By 18 months 57% of

the SCM group but only 27% of the MBT group had shown severe parasuicidal behavior, which corresponds to a relative risk reduction of 0.46 and an NNT of three. Breaking this down into suicide attempts, life threatening self-harm and hospital admission separately, we find significant benefits for suicide (NNT of 4), hospitalization (NNT of 6) and a somewhat less convincing effect for self harm (NNT of 5 but CI of 2-32). At the end of the 18 months was over 70% in the MBT group but only 45% in the SCM group were no longer taking psychoactive medication.

I don't have time to recount all the results but self-report measures of symptom severity, depression, social adjustment and interpersonal problems as well as blind ratings of general adaptive functioning (GAF) scores suggested that MBT benefitted patients substantially more than SCM.

However, an important moral of this tale for us was less how well the knights wearing MBT colours did and more how well both groups did on almost all the measures. Whilst our treatment may have worked better, both structured integrated and focused treatment protocols achieved better results than we might expect from spontaneous remissions of symptoms from follow-along studies. So, while the trial supported MBT, it also supported structured treatment approaches in general. The general utility of focusing on the patient's mind is supported and the trial strongly contraindicated the adoption of non-focused generic approaches

and the premature exclusive adaptation of any one model. We shall return to this theme later on.

[Page 31 →]

The story of the demise of the randomized controlled trial

I think it is now time to recount the stories of the sect of zealots who have tried to rule the kingdom of psychotherapy over the last couple of decades (Sackett, Richardson, Rosenberg, & Haynes, 2000; Strauss, Richardson, Glasziou, & Haynes, 2005). They wear many crests on their shields, but all swear allegiance to evidence based medicine as "the conscientious explicit and judicious use of current best evidence in making decisions about the care of individual patients" (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996). Their slogans are motivational, persuasive and essentialist rather than reportive, stipulative and operational.

The initiative is basically synonymous with attempts experimentally to establish a causal relationship between treatment and outcome. Hence their most highly valued strategy is the randomized controlled trial. Now, a strange thing happened last year. A high priest of evidence based medicine, none other than Sir Michael Rawlins, the Director of the National Institute of Health and Clinical Excellence (NICE), gave a Harveian Oration at the Royal College of Physicians about the over-valuation of randomized controlled trials in evidence-based

medicine (Rawlins, 2008). He pointed out that frequently such trials are inappropriate. There can be bioethical and legal problems in randomizing people to ineffective or harmful treatments. Some conditions are so rare and some treatment effects so massive that few would consider RCTs to be sensible. For example, parachutes are very widely used despite the fact that they have not been subjected to RCTs. He also pointed out that the null hypothesis of RCTs was there often for show rather than genuine conviction. It was certainly inappropriate where previous studies had already shown an effect and statistically and conceptually difficult if the aim of the study was to show no difference between treatment arms. Rawlins also identified some challenges in relation to applying theories of probability in RCTs.

Many studies use multiple variables to measure outcome, and this has massive potential to create confusion. With Anna Higgitt we have made a study of reviews of the value of SSRIs for paediatric depression (Fonagy & Higgitt, 2009). There are only about 15 RCTs but nearly 100 reviews have been published since 2005. These review the same investigations but come to dramatically different conclusions by focusing on different aspects of the results reported in the original investigations. The conclusions vary anywhere between 'ban all SSRIs' to 'use SSRIs as first line of treatment' with various gradations in between. ('Ban some SSRIs', 'use only one in first line treatment', 'use as second line treatment after psychological therapy has failed', etc.)

Sir Michael went on to revisit the issue most frequently alluded to in considering the appropriateness of RCTs in the Kingdom of Psychotherapy: how generalisable are the results of RCTs? Certainly, the settings in which psychotherapy RCTs take place are quite different from the real clinical situation

[Page 32 →]

(La Greca, Silverman, & Lochman, 2009; Weiss, Guidi, & Fava, 2009). Most are undertaken in selected populations for a finite time, rather than a heterogeneous population with many life problems and co-morbidities which might exclude them or they would exclude themselves. Certainly patients in trials and 'real life' differ in age, gender, severity, risk factors, co-morbidities, ethnicity, socio-economic status. As if that was not bad enough, the treatment given in psychotherapy RCTs rarely fits clinical reality in terms of dose (frequency of therapy), timing of administration, duration of therapy, inter-current treatments, the skills and commitment of the practitioners. There are major differences in the setting, the way the treaters are reimbursed, and their professional priorities – publication versus clinical care. There is a real question about whether the assessment of benefit obtained from a trial can be applied to ordinary clinical settings.

Sir Michael also drew attention to the assessment of harm. RCTs are appalling at testing for the possibility of harm (e.g. Lilienfeld, 2007; Roback, 2000). Jeffreys et al (Jefferys, Leakey, Lewis, Payne, & Rawlins, 1998) noted that

surveying drugs introduced between 1972 and 1994, they could find 22 that were withdrawn for safety reasons but only one withdrawn for lack of efficacy. This is hardly surprising, since RCTs are rarely powered up to scrutinize adverse events. The controversy over SSRIs and paediatric suicide is a good example. None of the placebo-controlled trials are large enough to show a significant difference in rates of spontaneous adverse events (SAEs). When taken together, they showed such events to be twice as frequent amongst children and adolescents taking active medication rather than placebo. But then those advocating for the use of these drugs argue that the study designs and samples are too heterogeneous to permit that kind of integration.

So perhaps, as Rawlins concluded, the knights of RCTs have been inappropriately elevated above those conducting observational studies. The hierarchy is illusory. Both kinds of study have advantages and disadvantages. In making decisions about cost-effective treatments against a background of limited resources, we need to appraise all the evidence and *exercise judgment*. RCTs are enormously costly. 153 pharmaceutical RCTs had median costs of £3,200,00, with an interquartile range of £2m - £6.25m.

RCTs are not only resource intensive in terms of money – they also take up considerable time and energy. Perhaps it is for this reason that their reporting is so often troubled by bias, aiming to emphasize the differences and be relatively silent about null findings. Take just one recent example on therapy for treatment

resistant depression in adolescents (Brent et al., 2008). Reading the abstract of this study we learn that when a young person did not respond to at least one SSRI, 40% are likely to improve if offered a medication switch. But they will improve even more if this switch is accompanied by CBT. The response rate increase is 15%. It makes no difference if they are switched to another SSRI or venlafaxine, but adding a psychological treatment clearly improves response rate. I don't want to bother you with the detail of this very complicated study other than to say that

[Page 33 →]

a disturbingly large number of young people withdrew from each arm of the treatment (out of 334 randomised, 102 – almost 1/3 – withdrew from the treatment, a significant number (41) because of adverse events). This raises real concerns about the acceptability of any of the treatment arms. Looking at responders, it seems clear that adding CBT improved response rates, although this seems less clear when we look at only those young people who completed the treatment trial (i.e. had the full benefit of CBT).

The story here is about something else, far more serious in relation to EBM. It is about three figures which I drew on the basis of numbers reported in very large, hard to access tables in the journal. These concerned children's self report of depression, suicidal ideation and the global assessment of independent

observers. I challenge you to see any difference whatsoever between the severity of depression, suicidal ideation or even the probably not so blind assessors' rating of global functioning. The conclusion reported in JAMA is that the combination of CBT and a switch to another antidepressant results in 'a higher rate of clinical response than did medication switch alone'. In the body of the paper, which is all about the value of CBT for this group, there is a telling sentence: "There were no differential treatment effects on scalar measures of depression, suicidal ideation, and functioning, nor were there treatment effects on suicide attempts, or self-harm-related measures"

I could bring many other similar examples of biased reporting of results. They are more common in the pharmacological literature, where commercial interests are great, but they are by no means unique to them. Nor would I wish to suggest that psychodynamic researchers are any less hampered by self serving biases than those from other orientations. Fifteen years ago one of our champion researchers, Lester Luborsky (1999) published an entertaining paper that showed you could magically divine the conclusion of a paper on psychotherapeutic outcome just by knowing the theoretical orientation of its first author. My point here is that notwithstanding the rhetoric of evidence based medicine, conscientious, explicit and judicious it frequently is not. As psychoanalysts we understand about how the mind has ways of bending reality to that which maximizes pleasure and minimizes psychic pain. Why should those

energetically in pursuit of making conscientious explicit and judicious use of current evidence be any different from the rest of us?

I don't have time to tell you the heroic tale of those who forced drug companies to reveal trial data indicating the risks of SSRIs, venlafaxine, mirtazapine etc. Let it suffice to say that when five years ago we published data including unpublished studies along with published ones on the efficacy of SSRIs for paediatric populations we found a disappointing reduction in effectiveness (Whittington et al., 2004). It took another four years before someone replicated our design for adult trial data (Turner, Matthews, Linardatos, Tell, & Rosenthal, 2008). 37 of 38 studies viewed by the FDA as having positive results were published. Out of 34 studies viewed by the FDA as having negative

[Page 34 →]

or questionable results, 22 were not published and 11 were published in a way that implied a positive outcome. Conscientious? Explicit? Judicious?

So, let's turn over a metaphoric page. The moral has to be: "read carefully, don't believe everything you do read and look for things that could and should be there".

A new intellectual framework for psychoanalytic psychotherapy research

This brings me to the last of my tales. This is a story of looking back towards the future - the title of a brilliant paper by my friends and colleagues Patrick Luyten and Sidney Blatt (Luyten & Blatt, 2007).

There are a number of episodes to this story and the first concerns the question 'how do we know if we have won?'

It turns out that our brave knights and their faithful researcher companions were competing in trials where not only the rules of engagement but also the criteria for winning had been defined by the Barons of Big Pharma. The outcome measures in the field of psychotherapy are self-report measures designed to be reactive to changes which neuro-chemical interventions tend to bring, by and large the blunting of awareness of distress caused by symptoms. Firstly, in many instances psychodynamic psychotherapy tries to increase awareness of distress rather than reduce it. Second, it is a devastating indictment of the entire system that there has been almost no client participation in defining outcome measures and the entire scheme is an edifice to evidence based practice as prescribed by professionals (Dolan, Lee, King, & Metclalfe, 2009). From a professional's standpoint as from that of the ordinary member of the public, physical role limitation, physical function and pain have high priority, whilst those suffering disorders rate dignity and general wellbeing (mood, global assessment of life, having a partner, job, lots of social contact) as more important. Wellbeing should

feature at least alongside, if not in place of, lists of symptoms in outcome studies (Pressman & Cohen, 2005).

Measures for the most part are arbitrary, measuring subtle psychological processes on arbitrary scales. Measures are arbitrary, but we reify them, we treat them and think of them as if they were not (Kazdin, 2006). A review of 2,000 RCTs in schizophrenia identified 640 scales, many of which were devised for the particular RCT and had no supporting data for validity or reliability. Unvalidated scales were more likely to show significant treatment effects than established scales. Arbitrary or not, our measures should be neutral in relation to the nature of treatment they intend to evaluate, otherwise we might find treatments targeting the scales of measurement rather than the disease process, which of course would be a travesty. For a number of years we have had non-reactive functional brain imaging measures of outcome available. Since Eisenberger and colleagues (Eisenberger, Lieberman, & Williams, 2003) demonstrated that social exclusion could activate the very same brain areas (anterior cingulated cortex, right ventral pre-frontal cortex) as the experience of physical pain, there have been literally hundreds of demonstrations

[Page 35 →]

of fMRI yielding accurate sensitive information related to subjective states. There have been 27 neuro-imaging studies of psychotherapy using a number of

imaging modalities, a range of diagnoses and therapeutic approaches (Carrig, Kolden, & Strauman, 2009).

These studies have their limitations, and almost no studies provide data on changes whilst treatment is going on. However, with a little ingenuity from the beekeepers and researcher bees, functional tasks could be designed which are specific not only to the disease condition but also to the hypothesized mechanism of action of the mode of therapy. Exploring the interplay of biological and psychological processes has the potential to enhance our understanding of the mode of action of psychotherapy. Multiple lines of evidence are likely to be needed to identify the mechanisms critical to particular types of intervention (Kazdin, 2008). The aim of this would not be to make psychological accounts redundant by providing a biological explanation, but rather to be more specific about what makes therapy work and to identify instances when it works well and for the long term. Biomarkers of change offer a way to dig deeper than information on symptom distress. Furthermore, basic biological research may suggest antecedents and primary pathologies that could prove to be targets for psychotherapeutic intervention. For example, neuro-biological researchers have identified impulsivity as a key target for preventative intervention for problems of addiction (Lejuez et al., 2007).

This brings us to one of the key questions facing our valiant knights as they look back into the future. The system of feudal patronage without primogeniture

which dominates our psychotherapeutic kingdom has led to fragmentation; once vast fiefdoms have been reduced to literally hundreds of pocket-sized plots of land. There are hundreds of therapies, many of them even possessing 'title-deeds' (evidence) supporting their potential to produce change. But are there really as many mechanisms for therapeutic change as there are modalities and orientations within modalities? Understanding how psychotherapy leads to change could group these together.

Are we as good at bringing about change as we think? Have all these knightly trials of therapy improved our effectiveness as therapists? Let me show you just one example. These are the effect sizes obtained in trials of cognitive behaviour therapy from 27 trials in youth depression. It should be obvious that rather than increasing in effect, the size of therapeutic benefit has been decreasing with time (Weisz, McCarty, & Valeri, 2006). Why? The obvious answer is that the Templar Knights of therapy have given the wrong instructions to their researcher bees. Using increasingly stringent criteria to define a therapy that works (for example increasingly realistic tests of effectiveness) will not yield a more efficacious treatment.

[Page 36 →]

By contrast, if we know more about how a change comes about, we might be able to identify alternative superior strategies that are more efficient in triggering

a change process. In particular, as we look back to think about how the gap between clinical effectiveness observed under research conditions may be effectively translated to ordinary daily practice, knowing about the mechanisms of change may help us selectively to guard features that by should not be diluted whilst being more relaxed about others that contribute to research vigour but not therapeutic efficacy.

Understanding mechanisms better will also help us to pinpoint moderators of treatment effectiveness. As clinicians we know that even the best approach doesn't work for everyone. But other than a few rules of thumb or truths passed down from our teachers, at the start of a treatment we have few ideas about what the indicators might be that suggest whether it is likely to work. Inexperienced therapists may often be as or more effective with certain chronic patients (e.g. Brown, Lambert, Jones, & Minami, 2005) because they do not have the experience that would tell them that their effort is likely to be futile.

Finally, the effectiveness research of the past 50 years, particularly the randomized controlled trials, have shown us that psychotherapy is causal in bringing about change. But demonstrating causation is but an illusion of explanation and a pernicious illusion at that. It gives rise to superstitious behaviour like Skinner's pigeons; rather than understanding how our treatments work, we merely repeat exactly the behaviors that led to the positive outcomes. In the UK we are involved in a nationwide trial of multi-systemic therapy

(Henggeler, Clingempeel, Brondino, & Pickrel, 2002) where enormous financial and human effort is being expended to replicate the intervention exactly as it was carried out in the United States at the Medical University of South Carolina. This is a futile and wasteful effort which is necessary only because of our ignorance of mechanisms of therapeutic action.

However, not only experimental studies but also observational studies are misleading in terms of identifying the effective components of treatments. There is the legend of the therapeutic alliance, still frequently taught in (k)night schools. It is often claimed that the therapeutic alliance is a mediator and mechanism of therapeutic change since the stronger the alliance, the greater the change observed (Klein et al., 2003). Correlational studies also show that alliance at the beginning of treatment predicts improvement in symptoms at the end (Cloitre, Chase Stovall-McClough, Miranda, & Chemtob, 2004). Do we need any more evidence to prove that it is the good relationship with the therapist that cures? More recent research that contrasted the outcome of patients with a number of therapists found that indeed differences between the effectiveness of therapists could be predicted by the strength of alliance they were likely to form with their patients (Baldwin, Wampold, & Imel, 2007) but differences in outcome between patients with the same therapist were unrelated to therapeutic alliance.

[Page 37 →]

If therapeutic alliance was the mechanism of change, then I would expect to do better with patients with whom I form a good alliance than those with whom my alliance is relatively poor. This turns out spectacularly not to be the case. So the ability to form an alliance does mark out our more talented therapists but what it is that they do more or less of that makes them more or less effective still remains a mystery.

Understanding why therapy works will increasingly require an understanding of the moderators of therapeutic effectiveness. Rapidly advancing biological research is providing increasingly persuasive evidence that there may be genetic limitations on how well therapy can work. Freud in his last major paper 'Analysis Terminable and Interminable' seemed to be acutely aware of the limitations of his technique, although he might understandably have misperceived some of the processes involved. Six years ago, in a widely quoted study, Avshalom Caspi and Terri Moffitt showed that the association between the number of stressful life events an individual experienced between 21 and 26 years and the probability of depression, suicidal ideation and suicide attempts was moderated by the 5HTT genotype (Caspi et al., 2003). Only those who had two of the short alleles of this genotype were likely to respond to four life events with increased suicidal ideation. The association between life events and suicidal ideation of those with two long alleles was completely absent.

This area of research has become a minor cottage industry, although many geneticists are appropriately skeptical about it (Risch et al., 2009). The most challenging finding, to me as an attachment theorist, is the report from Kochanska's laboratory which demonstrated that maternal sensitivity predicted infant security of attachment as it is supposed to only in infants with the short allele of the 5HTT genotype (Barry, Kochanska, & Philibert, 2008). Infants with the long allele were equally likely to be secure regardless of maternal sensitivity. Along similar lines but with older children, Kaufman reported an elegant study showing that the depressogenic effects of maltreatment could be mitigated by social support in individuals with the short allele of the 5HTT gene (Kaufman et al., 2004).

The hypothesis which cries out to be tested is that individuals whose depression is more likely to be associated with life events, as their serotonin transported gene marks them out to be environmentally sensitive, are also more likely to benefit from an environmental intervention such as psychotherapy. There is evidence indicative of the appropriateness of this way of thinking from those working on an attachment-based intervention to promote positive parenting and sensitive discipline. In this study, where parents of toddlers with externalizing behavioural problems were given video-feedback and other attachment theory guided interventions, the children benefitted if they had the 7-repeat allele on the DRD4 gene (Bakermans-Kranenburg, Van, Pijlman, Mesman, & Juffer, 2008).

The moral here is not that psychotherapy should not be offered to people without this or that allele, but rather that the mechanism by which therapy

[Page 38 →]

achieves its effect may be quite different for these constitutionally distinguishable groups of individuals. If we choose to ignore the reality of these differences in our clinical work, future generations are likely to judge this decision as unethical and unjustifiable and potentially as an indication of self serving attitude.

Conclusion

The moral of this long tale is that science is good for practice. Research is there not simply to defend the boundaries of our existing domains, but to help us deliver the forms of care which are best for our patients. To do this we have to understand better what causal mechanisms play a role in achieving patient benefit and also what circumstances can interfere with a treatment working. However, the flow of information should not be one way. Science, particularly neuroscience, will give us better ideas about how we can help our patients in more differentiated ways as it evolves.

But practice is also excellent for science. Practice has to tell researchers where knowledge is most needed and to ensure that science is firmly grounded in

everyday clinical care. Best evidence is only meaningful if used in proper argumentation. Argumentation is only meaningful if based on the best evidence in its building blocks. Jules Henri Poincaré wrote: “science is built up with facts as a house is with stone, but a collection of facts is no more a science than a heap of stones is a house”.

Thus ends my story of research in the land of psychotherapy. But I have been telling this story for long enough to know that arguments are persuasive only when they reach those parts of the brain where emotional significance is stored. So here is an observational study with personal significance for all in this room. Jeffery (2001) collected mortality rates for three groups of white men as a function of age. The sobering cumulative curve shows a positively accelerating trend for white males after the age of fifty. Medical training clearly impacts with a strong deceleration, but when we plot mortality rate of psychoanalysts along the same axis it is self evident that psychoanalysis is highly efficacious and recommends itself as a treatment of choice for life!

References

Bachar, E., Latzer, Y., Kreitler, S., & Berry, E. M. (1999). Empirical comparison of two psychological therapies. Self psychology and cognitive orientation in the treatment of anorexia and bulimia. *J Psychother Pract Res, 8*(2), 115-128.

Bakermans-Kranenburg, M. J., Van, I. M. H., Pijlman, F. T., Mesman, J., & Juffer, F. (2008). Experimental evidence for differential susceptibility: dopamine D4 receptor polymorphism (DRD4 VNTR) moderates intervention effects on toddlers' externalizing behavior in a randomized controlled trial. *Dev Psychol, 44*(1), 293-300.

Baldwin, S. A., Wampold, B. E., & Imel, Z. E. (2007). Untangling the alliance-outcome correlation: exploring the relative importance of therapist and patient variability in the alliance. *J Consult Clin Psychol, 75*(6), 842-852.

Barry, R. A., Kochanska, G., & Philibert, R. A. (2008). G x E interaction in the organization of attachment: mothers' responsiveness as a moderator of children's genotypes. *J Child Psychol Psychiatry, 49*(12), 1313-1320.

Bateman, A., & Fonagy, P. (2008). 8-year follow-up of patients treated for borderline personality disorder: mentalization-based treatment versus treatment as usual. *Am J Psychiatry, 165*(5), 631-638.

Bateman, A. W., & Fonagy, P. (1999). The effectiveness of partial hospitalization in the treatment of borderline personality disorder - a randomised controlled trial. *American Journal of Psychiatry, 156*, 1563-1569.

Bateman, A. W., & Fonagy, P. (2006). *Mentalization Based Treatment for Borderline Personality Disorder: A Practical Guide*. Oxford: Oxford University Press.

Bateman, A. W., & Fonagy, P. (in press). 8-year follow-up of patients treated for borderline personality disorder - mentalization based treatment versus treatment as usual. *American Journal of Psychiatry*.

Beck, A. T., & Bhar, S. S. (2009). Analyzing effectiveness of long-term psychodynamic psychotherapy. *Jama, 301*(9), 931; author reply 932-933.

Beutel, M., Rasting, M., Stuhr, U., Ruger, B., & Leuzinger-Bohleber, M. (2004). Assessing the Impact of Psychoanalyses and Long-Term Psychoanalytic Therapies on Health Care Utilization and Costs. *Psychotherapy Research, 14*, 146-160.

Blum, N., St John, D., Pfohl, B., Stuart, S., McCormick, B., Allen, J., et al. (2008). Systems Training for Emotional Predictability and Problem Solving (STEPPS) for outpatients with borderline personality disorder: a randomized controlled trial and 1-year follow-up. *Am J Psychiatry, 165*(4), 468-478.

Brent, D., Emslie, G., Clarke, G., Wagner, K. D., Asarnow, J. R., Keller, M., et al. (2008). Switching to another SSRI or to venlafaxine with or without cognitive behavioral therapy for adolescents with SSRI-resistant depression: the TORDIA randomized controlled trial. *Jama, 299*(8), 901-913.

Brown, G. S., Lambert, M. J., Jones, E. R., & Minami, T. (2005). Identifying highly effective psychotherapists in a managed care environment. *Am J Manag Care, 11*(8), 513-520.

Carrig, M. M., Kolden, G. G., & Strauman, T. J. (2009). Using functional magnetic resonance imaging in psychotherapy research: A brief introduction to concepts, methods, and task selection. *Psychother Res*, 1-9.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., et al. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science, 301*(5631), 386-389.

Chanen, A. M., Jackson, H. J., McCutcheon, L. K., Jovev, M., Dudgeon, P., Yuen, H. P., et al. (2008). Early intervention for adolescents with borderline personality disorder using cognitive analytic therapy: randomised controlled trial. *Br J Psychiatry, 193*(6), 477-484.

Clarkin, J., Levy, K. N., Lenzenweger, M. F., & Kernberg, O. F. (2007). Evaluating three treatments for borderline personality disorder: A multiwave study. *American Journal of Psychiatry, 164*, 922-928.

Cloitre, M., Chase Stovall-McClough, K., Miranda, R., & Chemtob, C. M. (2004). Therapeutic alliance, negative mood regulation, and treatment outcome in child abuse-related posttraumatic stress disorder. *J Consult Clin Psychol, 72*(3), 411-416.

Cohen, J. (1962). The statistical power of abnormal-social psychological reserach: a review. *Journal of Abnormal and Social Psychology, 65*, 145-153.

Cottraux, J., Note, I. D., Boutitie, F., Milliery, M., Genouihlac, V., Yao, S. N., et al. (2009). Cognitive therapy versus Rogerian supportive therapy in borderline personality disorder. Two-year follow-up of a controlled pilot study. *Psychother Psychosom, 78*(5), 307-316.

Cuijpers, P., van Straten, A., van Oppen, P., & Andersson, G. (2008). Are psychological and pharmacologic interventions equally effective in the treatment of adult depressive disorders? A meta-analysis of comparative studies. *J Clin Psychiatry, 69*(11), 1675-1685; quiz 1839-1641.

Dare, C., Eisler, I., Russell, G., Treasure, J., & Dodge, L. (2001). Psychological therapies for adults with anorexia nervosa: randomised controlled trial of out-patient treatments. *Br J Psychiatry, 178*, 216-221.

de Maat, S., de Jonghe, F., Schoevers, R., & Dekker, J. (2009). The effectiveness of long-term psychoanalytic therapy: a systematic review of empirical studies. *Harv Rev Psychiatry, 17*(1), 1-23.

Dolan, P., Lee, H. J., King, D., & Metclalfe, R. (2009). How does NICE value health? *BMJ, 339*, 371-373.

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An FMRI study of social exclusion. *Science, 302*(5643), 290-292.

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

Fonagy, P., & Higgitt, A. (2009). *SSRIs should they be the first line treatment for pediatric depression or should they be banned: A systematic reveiw of not so systematic reviews of the SSRI controversy*.Unpublished manuscript.

Giesen-Bloo, J., van Dyck, R., Spinhoven, P., van Tilburg, W., Dirksen, C., van Asselt, T., et al. (2006). Outpatient psychotherapy for borderline personality disorder: randomized trial of schema-focused therapy vs transference-focused psychotherapy. *Arch Gen Psychiatry, 63*(6), 649-658.

Glass, R. M. (2008). Psychodynamic psychotherapy and research evidence: Bambi survives Godzilla? *Jama, 300*(13), 1587-1589.

Gregory, R. J., Virk, S., Chlebowski, S., Kang, D., Remen, A. L., Soderberg, M. G., et al. (2008). A controlled trial of psychodynamic psychotherapy for co-occurring borderline personality disorder and alcohol use disorder. *Psychotherapy: Theory, Research, Practice, Training, 45*(1), 28–41.

Henggeler, S. W., Clingempeel, W. G., Brondino, M. J., & Pickrel, S. G. (2002). Four-year follow-up of multisystemic therapy with substance-abusing and substance-dependent juvenile offenders. *J Am Acad Child Adolesc Psychiatry, 41*(7), 868-874.

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. Special Issue: Psychotherapy research. *American Psychologist, 41*, 159-164.

Huber, D., Denscherz, C., Gastner, J., Henrich, G., & Klug, G. (submitted). Psychodynamic long-term psychotherapies and cognitive-behavior therapy in comparison.

Jeffery, E. H. (2001). The mortality of psychoanalysts. *Journal of the American Psychoanalytic Association, 49*, 103-111.

Jefferys, D. B., Leakey, D., Lewis, J. A., Payne, S., & Rawlins, M. D. (1998). New active substances authorised in the United Kingdom between 1972 and 1994. *British Journal of Clinical Pharmacology, 45*, 151–156.

Kaufman, J., Yang, B. Z., Douglas-Palumberi, H., Houshyar, S., Lipschitz, D., Krystal, J. H., et al. (2004). Social supports and serotonin transporter gene moderate depression in maltreated children. *Proc Natl Acad Sci U S A, 101*(49), 17316-17321.

Kazdin, A. E. (2006). Arbitrary metrics: implications for identifying evidence-based treatments. *Am Psychol, 61*(1), 42-49; discussion 62-71.

Kazdin, A. E. (2008). Understanding how and why psychotherapy leads to change. *Psychother Res*, 1-11.

Klein, D. N., Schwartz, J. E., Santiago, N. J., Vivian, D., Vocisano, C., Castonguay, L. G., et al. (2003). Therapeutic alliance in depression treatment: controlling for prior change and patient characteristics. *J Consult Clin Psychol, 71*(6), 997-1006.

Klug, G., & Huber, D. (2009). Psychic structure: exploring an empirically still unknown territory. *J Am Psychoanal Assoc, 57*(1), 149-173.

Knekt, P., Lindfors, O., Harkanen, T., Valikoski, M., Virtala, E., Laaksonen, M. A., et al. (2008). Randomized trial on the effectiveness of long-and short-term psychodynamic psychotherapy and solution-focused therapy on psychiatric symptoms during a 3-year follow-up. *Psychol Med, 38*(5), 689-703.

Kopta, S. M., Lueger, R. J., Saunders, S. M., & Howard, K. I. (1999). Individual psychotherapy outcome and process research: Challenges leading to greater turmoil or a positive transition? *Annual Review of Psychology, 50*, 441-469.

Korner, A., Gerull, F., Meares, R., & Stevenson, J. (2006). Borderline personality disorder treated with the conversational model: a replication study. *Comprehensive Psychiatry, 47*, 406-411.

Kriston, L., Holzel, L., & Harter, M. (2009). Analyzing effectiveness of long-term psychodynamic psychotherapy. *Jama, 301*(9), 930-931; author reply 932-933.

La Greca, A. M., Silverman, W. K., & Lochman, J. E. (2009). Moving beyond efficacy and effectiveness in child and adolescent intervention research. *J Consult Clin Psychol, 77*(3), 373-382.

Laupacis, A., Sackett, D. L., & Roberts, R. S. (1988). An Assessment of Clinically Useful Measures of the Consequences of Treatment. *N Engl J Med, 318*(26), 1728-1733.

Leichsenring, F., & Rabung, S. (2008). Effectiveness of long-term psychodynamic psychotherapy: a meta-analysis. *Jama, 300*(13), 1551-1565.

Leichsenring, F., & Rabung, S. (2009). Author's reply. *Jama, 301*, 932-933.

Lejuez, C. W., Aklin, W., Daughters, S., Zvolensky, M., Kahler, C., & Gwadz, M. (2007). Reliability and validity of the youth version of the Balloon Analogue Risk Task (BART-Y) in the assessment of risk-taking behavior among inner-city adolescents. *J Clin Child Adolesc Psychol, 36*(1), 106-111.

Leuzinger-Bohleber, M., Stuhr, U., Ruger, B., & Beutel, M. (2003). How to study the 'quality of psychoanalytic treatments' and their long-term effects on patients' well-being: a representative, multi-perspective follow-up study. *Int J Psychoanal, 84*(Pt 2), 263-290.

Lilienfeld, S. O. (2007). Psychological treatmetns that cause harm. *Perspectives on Psychological Science, 2*, 53-70.

Linehan, M. M., Comtois, K. A., Murray, A. M., Brown, M. Z., Gallop, R. J., Heard, H. L., et al. (2006). Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. *Arch Gen Psychiatry, 63*(7), 757-766.

Luborsky, L., Diguer, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., et al. (1999). The researcher's own therapy allegiances: A 'wild card' in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice, 6*, 95-106.

Luyten, P., & Blatt, S. J. (2007). Looking Back Towards the Future: Is It Time to Change the DSM Approach to Psychiatric Disorders? The Case of Depression. *Psychiatry, 70*(2), 85-99.

Maina, G., Rosso, G., & Bogetto, F. (2009). Brief dynamic therapy combined with pharmacotherapy in the treatment of major depressive disorder: long-term results. *J Affect Disord, 114*(1-3), 200-207.

Pressman, S. D., & Cohen, S. (2005). Does positive affect influence health? *Psychol Bull, 131*(6), 925-971.

Puschner, B., Kraft, S., Kachele, H., & Kordy, H. (2007). Course of improvement over 2 years in psychoanalytic and psychodynamic outpatient psychotherapy. *Psychology and Psychotherapy: Theory, Research and Practice, 80*(1), 51-68.

Rawlins, M. (2008). *De Testimonio: On the evidence for decisions about the use of therapeutic interventions*. London: Royal College of Physicians (The Harveian Oration).

Risch, N., Herrell, R., Lehner, T., Liang, K. Y., Eaves, L., Hoh, J., et al. (2009). Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. *Jama, 301*(23), 2462-2471.

Roback, H. B. (2000). Adverse outcomes in group psychotherapy: risk factors, prevention, and research directions. *J Psychother Pract Res, 9*(3), 113-122.

Roepke, S., & Renneberg, B. (2009). Analyzing effectiveness of long-term psychodynamic psychotherapy. *Jama, 301*(9), 931-932; author reply 932-933.

Sackett, D. L., Richardson, W. S., Rosenberg, W. M., & Haynes, R. B. (2000). *Evidence based medicine: How to practice and teach EBM* (2nd ed.). New York & Edinburgh: Churchill Livingstone.

Sackett, D. L., Rosenberg, W. M., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal, 312*, 71-72.

Strauss, S. E., Richardson, W. S., Glasziou, P., & Haynes, R. B. (2005). *Evidence based medicine: How to practice and teach EBM* (3rd ed.). New York: Elsevier.

Svartberg, M., Stiles, T. C., & Seltzer, M. H. (2004). Randomized, controlled trial of the effectiveness of short-term dynamic psychotherapy and cognitive therapy for cluster C personality disorders. *American Journal of Psychiatry, 161*, 810-817.

Thombs, B. D., Bassel, M., & Jewett, L. R. (2009). Analyzing effectiveness of long-term psychodynamic psychotherapy. *Jama, 301*(9), 930; author reply 932-933.

Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med, 358*(3), 252-260.

Tyrer, P., Tom, B., Byford, S., Schmidt, U., Jones, V., Davidson, K., et al. (2004). Differential effects of manual assisted cognitive behavior therapy in the treatment of recurrent deliberate self-harm and personality disturbance: the POPMACT study. *J Personal Disord, 18*(1), 102-116.

Wampold, B. E. (2001). *The Great Psychotherapy Debate : Models, Methods, and Findings* Hillsdale, NJ: Laurence Erlbaum Associates.

Wampold, B. E., Imel, Z. E., & Minami, T. (2007). The story of placebo effects in medicine: evidence in context. *J Clin Psychol, 63*(4), 379-390.

Weiss, A. P., Guidi, J., & Fava, M. (2009). Closing the efficacy-effectiveness gap: translating both the what and the how from randomized controlled trials to clinical practice. *J Clin Psychiatry, 70*(4), 446-449.

Weisz, J. R., McCarty, C. A., & Valeri, S. M. (2006). Effects of psychotherapy for depression in children and adolescents: a meta-analysis. *Psychol Bull, 132*(1), 132-149.

Westen, D., & Bradley, R. (2005). Empirically supported complexity: Rethinking evidence-based practice in psychotherapy. *Current Directions in Psychological Science, 14*(5), 266-271.

Whittington, C. J., Kendall, T., Fonagy, P., Cottrell, D., Cotgrove, A., & Boddington, E. (2004). Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. *Lancet, 363*(9418), 1341-1345.