Assessing and Managing Risk in Mental Health Services: The FACE Risk Profile

Dr Paul Clifford

Contents

List of Tables and Figures

Tables

Figures

The Author

Dr Paul Clifford is a clinical psychologist who has been involved in national research and development programmes in the UK for over 25 years. He has a background in mathematical logic, coding and classification and the semantics of natural language. His career includes: Programme Co-ordinator for the Sainsbury Centre for Mental Health, Director of Mental Health Resource Management systems development for the Department of Health, and first Director of the British Psychological Society's Centre for Outcomes Research and Effectiveness at UCL. He has developed over 100 assessment and outcome measurement tools, which have been used both nationally and internationally. He is currently CEO of Imosphere Ltd.

Please address any correspondence to: paul.clifford@imosphere.co.uk.

The Publishers

Imosphere Ltd is a specialist provider of health and social care assessment tools, data analysis and data capture systems. The company comprises a mix of senior health and social care practitioners, researchers with a strong informatics, health research and statistical background; leading edge system developers; and experienced project managers with a health and social care background. We have worked in this field for over 20 years and our goal is to improve people's lives by transforming the use of healthcare information. We do this by producing novel software tools and also clinical forms and assessment tools, many of which have been used on a national basis. We are currently working in the UK, the US, Ireland and mainland Europe.

The Risk Profile

Copies of the FACE Risk Profile are available from info@imosphere.co.uk.

Foreword

The assessment and management of risk in mental health services is a major focus of policy at present, as it has been for most of the last decade. This monograph describes both the development of a risk assessment tool for mental health services, the FACE Risk Profile; and its use to explore the nature of judgements and perceptions of risk. The Risk Profile is itself in part a response to policy imperatives. However, in presenting the tool, the thinking behind it and some data, this monograph also tries to identify some of the practical and technical barriers to effective risk assessment and management; to identify where there are opportunities for taking a different and perhaps more fertile approach; to clarify some of the conceptual issues; and to identify areas where the issues are ill-understood and (considerably) more evidence is required.

One of the worst aspects of modern 'blame culture' and the defensive approach to risk is that it tends to encourage premature closure – an implicit assumption that either we must already know most of what can be known or that resources do not provide any scope for us to enhance our knowledge – and so the best we can do is to review the evidence such as it is, pick out the key variables and build assessment tools around them. The result is a narrowness of focus: on process rather than on outcomes; on what could, should, or might have been done in an individual case rather than the management of risk across an organisation; on 'research' without full consideration of the meaning of the data; or on covering ones back rather than building the future.

In developing the FACE Risk Profile we tried to do something more than tick the right boxes: to develop something that is useful and can provide a solid basis for effectively managing risk. However, we also wanted to develop a tool that could generate new knowledge and understanding about risk, one which could help improve and inform clinical practice. The tool is now widely used across the UK, with many Trusts having used the tool for many years. This provides some evidence that we have to some extent achieved our first objective.

We did not expect the number of complex and sometimes surprising findings that the data uncovered. As well as confirming belief in the value of a 'learning health system' we hope that what we have found will inspire others, whether researchers or clinicians, to further investigate this important, fascinating and surprisingly unexplored area.

1 Risk Assessment in Mental Health

According to modern policy documents, 'risk assessment and risk management is at the heart of effective mental health practice' (Department of Health 2000 (p.23)). Common agreed risk assessment and risk management process 'is a feature of a truly integrated system of care'. A key component of a health and social assessment therefore needs to be a 'risk assessment', capable of recording risk in a standard form and able to support both the effective clinical management of risk and the communication of risk to the different professionals and agencies involved in delivering care. This chapter therefore examines the literature on assessment of risk in mental health services, with the emphasis on identifying methodological and content issues relating to the design of a tool for routine use.

The review itself is in four parts. The first two sections briefly review the literature on risk to self and risk to others in people with mental illness, focusing on the most extreme risks: risk of suicide and homicide. The third section focuses on the question of prediction. The fourth section reviews some of the risk assessment tools currently in use. Finally, the review is used to generate a set of requirements which should ideally be met by a risk assessment tool for use in routine practice.

1.1 Suicide and risk to self

Approximately 1000 service users per year in the UK commit suicide having been in contact with mental health services the week before death (Appleby et al 1999). There is good data on the characteristics of these individuals, collected by the National Confidential Inquiry into Suicide and Homicide by People with Mental Illness, which was established in 1992. The inquiry received notification of over 10,000 suicides and analysed data on about 2000 people in contact with mental health services in the year prior to suicide. About two thirds of the suicides in the mental health population were male and the median age was 41. Forty-one percent lived alone (ibid.).

The inquiry confirmed previous findings that high suicide rates in mental health service users are associated with acute episodes of illness, recent hospital discharge, social factors such as living alone, and clinical features such as drug or alcohol misuse and non-fatal self-harm. The two most common clinical diagnoses were affective disorder (including depression and bipolar) (42%), and schizophrenia (21%). Personality disorder (11%) and drug or alcohol dependence (13%) also featured. About half the service users had a clinical history going back >5 years. About two thirds had a previous history of self-harm, about one fifth had a history of violence, two fifths had a history of alcohol misuse and a quarter a history of drug misuse. About 25% had a history of non-compliance with care in the month before suicide (e.g. unplanned discharge) (Appleby et al 1999).

The data from the confidential inquiry is supplemented helpfully by a number of more clinically-oriented studies. Morgan and Stanton (1997) compared service users who had died through suicide either while receiving in-patient care or within two months of discharge from hospital, with a similar series reported 10 years previously. Clinicians' perceptions of service users' behaviour were compared with concurrent controls. Service users in the more recent study were younger, more often male, and a greater proportion had been discharged from in-patient status. Issues that complicated risk assessment

included short-lasting misleading clinical improvements, variability in degree of distress, and a reluctance to discuss suicidal ideas. Over a range of perceived behaviours it was not possible to distinguish suicides from controls. Suicidal behaviours, including the writing of suicide notes and specific plans, did not permit accurate distinction of those service users at risk of suicide. They therefore concluded that in assessing suicide risk paramount importance should be attached to monitoring suicidal ideation and addressing the hazards that might complicate this procedure.

In the American managed care context, Hall et al reviewed risk factors for suicide in 100 service users who made severe suicide attempts (Hall et al 1999). They found that severe anxiety, panic attacks, a depressed mood, a diagnosis of major affective disorder, recent loss of an interpersonal relationship, recent abuse of alcohol or illicit substances coupled with feelings of hopelessness, helplessness, worthlessness, global or partial insomnia, anhedonia, inability to maintain a job, and the recent onset of impulsive behaviour were 'excellent' predictors of suicidal behaviour. The presence of a specific suicide plan or suicide note was not, confirming Morgan and Stanton's earlier finding.

Stanton and Morgan also noted that short-term clinical improvements could provide a misleading guide to risk of suicide. This assertion is perhaps supported by the finding of Appleby et al that at final contact with mental health services, the risk of suicide was considered to have been either absent or low in 84% of cases by clinicians (p.1238). Both findings suggest the possibility that clinicians sometimes place too great an emphasis on current mental state as the key predictor of future behaviour. However, since most cases of suicide occur when risk is not apparent, the data could also be taken to indicate that where risk *is apparent,* clinicians are generally good at taking preventive action. At a practical level, it is plausible that such factors as the concealment of suicidal ideation by service users determined to act would result in risk not being apparent. Thus depending upon the point of view taken, the same data can be interpreted in opposite directions.

The literature thus identifies a range of demographic, historical, clinical, behavioural, social and service-related characteristics that appear to be associated with suicide, all of which would be natural candidates for inclusion within a clinical risk assessment tool. Assessment of suicide risk needs to place the correct emphasis upon, respectively, current factors and historical factors; and ideation as opposed to behaviours. Indeed, the potential value of accurate assessment is highlighted by the finding of the confidential inquiry that mental health teams regarded one in five suicides as preventable. In three fifths of cases, teams were able to specify at least one measure that could have reduced risk, most commonly measures to improve service user compliance (29%) and closer supervision (26%). It is unclear what weight to attach to such observations, however, because whilst looking retrospectively at a single case it might appear that it would have been relatively straightforward to take additional measures, across the whole clinical caseload being managed by the team at that time it might not have been realistic to undertake additional measures for all service users with similar presentation. In practice, clinicians tend to be very busy so additional resources are allocated based upon perceived clinical priority. Were cases with perceived low risk, and therefore low priority, to be provided with additional input then resources could be stretched past breaking point. More needs to be known, therefore, about the frequency of presentation of suicide risk in routine practice for an informed judgement to be made about how many cases could in fact be prevented.

Finally, suicide is the most extreme example of a wide range of behaviours which can result in self-harm, including both behaviours deliberately aimed at self-harm, such as parasuicide and self-cutting; behaviours of omission, such as self-neglect; and other potentially self-harmful behaviour such as substance misuse, eating disorders and promiscuity. Risk assessment cannot afford to ignore such behaviours. Thus although policy is clear on the need for risk assessment, and national policy also targets a reduction in suicide rates, it is far from clear that the two policies are complementary. One of the challenges of a risk assessment tool is therefore to cover a broad range of risks to self, whilst maintaining an adequate focus on suicide prevention.

1.2 Homicide and risk to others

Taylor and Gunn (1999) analysed the data on homicides between 1957 and 1995 and showed that the homicide rate involving people with a mental disorder since the policy of community care was implemented has remained constant at about 40 per year – a finding in marked contrast to the media attention given to the most distressing cases.

Good data on homicide cases involving people with mental illness derives from two sources. First, in 1994, The Department of Health decided that in cases of homicide an independent inquiry should always be held (NHS Executive 1994). The findings of such inquiries were published and are freely available. Munro and Rumgay (1997) analysed 40 cases of homicide covered by such enquiries and concluded that (according to the inquiry panels) 'improved risk assessment would have identified 11 of the cases and the homicides could have been prevented in only nine of them. Seventeen other deaths could have been prevented if professionals had responded more efficiently to signs that the service users were relapsing, although they gave no clear signs that their illness would include violent behaviour on this occasion'.

A second major data source is the confidential inquiry, which examined 718 homicide convictions occurring notified between April 1996 and November 1997 (Shaw et al 1999). The majority were young, male, single and unemployed. About two fifths were living alone. About one in seven of the perpetrators had been in contact with mental health services in their lifetime. The 71 individuals judged to be suffering from a mental disorder at the time of homicide were similar in most respects to the broader perpetrator population. They differed in being slightly younger (median age 27 rather than 34) and drugs or alcohol were less likely to have contributed to the offence. They were also considerably less likely to murder a stranger (7% compared to 25%). Immediate risk was thought to be absent or low at last contact in about nine in ten cases. Mental health teams regarded only 12% of cases as preventable, although in 40 cases measures were specified that could have reduced risk, notably measures to ensure better compliance with treatment.

The findings of the confidential inquiry and the conclusions from the national enquiries therefore present somewhat of a contrast: the latter finding a relatively small proportion of cases to be preventable whilst the former suggests that nearly two thirds (albeit of a very different type of sample) were preventable.

Homicides are, of course, at one end of a spectrum of behaviours that result in harm to others, ranging from relatively minor assaults to others resulting in serious harm but not death. The concept of 'risk to others' is therefore far broader than that of homicide. It follows that the aim of minimising risk to others is far broader than that of homicide

prevention. This distinction has practical implications since it cannot be assumed that measures designed to minimise risk to others will also reduce risk of homicide or vice versa. For example, given that resources are limited, it is conceivable that a broad focus on minimisation of risk to others could reduce incidents of harm to others whilst simultaneously increasing homicides by virtue of deflecting resources away from the most high-risk individuals.

A related issue is the question of how to conceptualise 'dangerousness'. The fact that some individuals are plainly a menace does not mean that 'dangerousness' is a general attribute of individuals, possessed in higher or lower doses. The Sinclair Report (Lingham, Candy and Bray, 1996) used the image of a 'ladder of dangerousness' up and down which a service user may move. This is a seductive image, but one which avoids key questions such as whether everyone moves up and down the same ladder and whether everyone's ladder includes a rung marked 'homicide'. Indeed, one of the problems with the concept of dangerousness is the implication of a continuum along which individuals can move or be placed. In practice, the likelihood of an incident involving harm to others may be a product of a complex interaction between social, clinical and environmental variables. A study by Swanson et al (1998) illustrates the complexity of risk to others in routine mental health services. They examined the determinants of violent behaviour in a sample of 331 adults with severe mental disorders in community-based treatment. They found that amongst lower functioning service users, higher frequency of social contact with family and friends was found to be significantly associated with risk of violence, whereas amongst the better functioning population, frequent social contact was associated with lower risk of violence and greater satisfaction with relationships. They therefore concluded that 'social contact may be a mixed blessing for SMI individuals. For some, it signals a positive quality of life, but for others--particularly those with extreme psychiatric impairment--frequent contact may add to conflict, stress, and increased potential and opportunity for physical violence. The impact of psychiatric impairment on violent behaviour cannot be known in isolation, but must be considered in a social context. Effective community-based strategies to anticipate and prevent violence in the lives of persons with severe mental illness must take into account such interactions between social and clinical variables'.

Notwithstanding such complexities, apparently in response to a number of high profile cases of homicide, over the past ten years national policy in the UK has shifted in the direction of attempting to focus on the targeting of those posing most risk – a trend which has been consistently criticised by professionals who have argued that resources would be better directed at improving the overall quality of care for all (Eastman 1997). The policy requirement of 'agreed risk assessment and risk management process' is conveniently silent as to which objective it is intended to meet: is it to minimise risk to others and/or to reduce preventable homicides by people with a mental illness? Is it assumed that these (potentially conflicting) objectives are one and the same or is the aim simply to be able to say 'we did our best' when the inevitable happens? Policy, as ever, appears to be a meld of all these motivations. At a practical level, paralleling the conclusion of the previous discussion on suicide and risk to self, one of the challenges of a risk assessment tool must be to reflect this complexity: to maintain the right balance between coverage of a broad range of behaviours resulting in risk to others whilst maintaining an adequate focus on homicide prevention – at the same time as attempting to ensure that even if things do go wrong, at least the basics were covered.

1.3 Prediction

It is natural to consider to what extent future risk behaviour, whether directed at self or others, may be predicted from knowledge of past and current behaviour. To date, the predictiveness of rating scales has been disappointing – which is either an argument for developing better scales or for a different approach, depending upon one's point of view. Menzies, Webster, McMain et al. (1994, p.25) evaluated the predictive validity of two rating scales with disappointing results, leading them to conclude that 'the objective of a standardised, reliable, generalisable set of criteria for dangerousness predictions, in law and in mental health, is still an elusive and distant objective'.

Empirical studies to improve predictiveness face many methodological difficulties, summarised by Monahan: 'Four methodological problems have especially plagued actuarial research on the assessment of the potential of the mentally disordered for violence: inadequacy of cues or factors chosen to forecast whether violence will occur, inability to determine the extent of violence within the population studied, limited applicability of research designs used to validate risk factors, and failure to coordinate research efforts in the field' (Monahan 1996, p.7). The most problematic of these is probably determining the extent of violence within the population studied, since derivation of an accurate estimate would require that intervention be withheld from people perceived to be at risk of violence in order to ascertain whether in fact the predicted violence occurs. Only thus would it be possible to assess with precision how predictive known risk factors are of future behaviour. This has obvious legal and ethical objections a corollary of which is that it may never be possible to know how many incidents are being prevented by the timely intervention of mental health professionals based upon their assessment of risk. This is possibly less of an obstacle in relation to suicide risk, given the known high base mortality rates amongst populations with certain diagnoses, such as schizophrenia (10% lifetime risk of suicide) and depression (15% lifetime risk of suicide) (Morgan and Owen 1990).

Even if such methodological difficulties were overcome, there remains the problem of demonstrating the utility of actuarial solutions in routine service contexts, as opposed to those contexts such as forensic services where in-depth risk assessment is a sine qua non of clinical practice. Obvious difficulties are the acceptability to clinicians of actuarial predictions on those occasions when they appear to be at variance with clinical judgement and the feasibility of implementing such procedures in routine practice given heavy workloads, variable staff skills mix and resource shortages. However, probably the key question is whether actuarial solutions can tell clinicians something they don't know already. Subjectively, service users may be divided into three groups: those who clearly do not pose a risk, those who clearly do pose a risk and those for whom the level of risk is unclear. For an actuarial solution to be of significant clinical utility it must be accurately predictive for the middle group, those where the clinical decision on risk is unclear[1]. Otherwise, all the actuarial tool will do is confirm what is already self-evident – an admirable attribute but unlikely to be compelling given the cost of routine use. In the same way, a weather forecasting service would be of limited interest if it was only able to accurately forecast rain when an imminent downpour was obvious to anyone looking out of the window or paying attention to everyday signs of change in the weather. What

---

[1] This might include service users where clinicians did not agree on perceived risk, service users for whom it is known that clinical judgements are poorly predictive (see McNeil et al 2000).

we want from the weather forecast is accurate prediction of that which we could not otherwise predict from informed observation. This is a tall order (at least as applied to risk) and it remains to be seen whether actuarial tools will ever achieve this goal.

Even if the dream of prediction is never fulfilled there are still important lessons to be learnt from the actuarial literature. One of the main arguments in favour of actuarial tools is that they are better predictors than clinical judgement, perhaps because they are less influenced by present circumstantial factors. McNeil and Binder (1995) evaluated the accuracy of physicians' assessment of likelihood that 226 in-service users would physically attack someone during the first week of admission. Service users with psychotic disorders and mania were most likely to be accurately (positively) predicted. They concluded that whilst clinicians can accurately classify the potential for violence in the majority of service users at admission, there were nevertheless systematic errors that characterised inaccurate risk assessment, such as over-emphasis on gender or ethnicity.

The latter finding has to be qualified by the later finding of McNeil et al (1998) that the greater the confidence of clinicians in their risk assessment, the greater the predictive accuracy of that assessment: 'when clinicians had a high degree of confidence, their evaluations of risk of violence were strongly associated with whether or not service users became violent. At moderate levels of confidence, clinicians' risk estimates had a lower, but still substantial relationship with the later occurrence of violence. However, when clinicians had low confidence, their assessments of potential for violence had little relationship to whether or not the service users became violent'. In similar vein, McNeil et al (2000) found that the higher the level of agreement between clinicians on risk, the more accurate the prediction.

In sum, few would argue with the conclusion of Allen (1997) that 'actuarial indicators need to be supported by structured thorough clinical investigation'. Similarly, Ferris et al (1997) summarising a comprehensive review of the MEDLINE and PSYCHINFO databases, concluded that 'research shows that risk assessments do have validity for use in short-term prediction and that it is possible to develop clinical guidelines in this area. A combined clinical and research approach holds the most promise for improving the accuracy of probability estimates, and most published guides and tools rely on such a combination'. 'Hedge your bets' would seem to be the recommended strategy.

1.4   Tools in use

This section briefly reviews a number of risk assessment tools widely referred to in the literature and also in routine use in at least some mental health settings. The focus is on issues related to potential routine use, rather than on whether they are useful in the more specialist forensic contexts from which they have generally emerged. The majority of the tools referred to also tend to focus on risk to others rather than risk to self.

One of the longest-established tools is Hare's Psychopathy Checklist (Hare 1980). Dolan and Doyle (2000) concluded that the tool appears to be a key predictor of violent recidivism in a variety of settings. However, Hare's checklist has a strong forensic emphasis, emphasising poor behavioural control, forensic history and pathological personality traits with a particular focus on psychopathy and narcissistic personality traits. Whilst it has obvious relevance to forensic psychiatric contexts it is inappropriate for conventional secondary mental health services.

In order to overcome the methodological deficiencies of previous actuarial approaches, Monahan et al embarked upon a series of studies resulting in the production of a tool designed for routine clinical use, the Iteration Classification Tree (Monahan et al (2000)). This claims to restrict the risk factors used to generate the actuarial tool to those 'capable of being routinely assessed in clinical practice'. It also uses more sophisticated statistical techniques than simple regression to generate the prediction. However, the authors warn that it can be time and resource intensive to administer. The reported algorithm classifies newly-discharged service users into three groups based on a set of 106 'clinically-feasible' risk factors, a number that is perhaps stretching the concept of clinical feasibility. The clinical utility of the measure is hard to evaluate. The three groups are low risk, high risk and unclassified. Of those placed in the high risk group, just over half committed a violent act in the next several months compared to about 5% of the low risk group and 25% of the unclassified group. Without knowing what proportion of the high risk group would have been classified as such by clinicians, it is difficult to assess what added value the tool is providing. Furthermore, the fact that a quarter of the unclassified group also committed acts of violence suggests that classification in this group is also a reasonable indicator of risk. Were clinicians to classify both 'high risk' and 'unclassified' groups as requiring special risk management measures then about half the total sample would fall into this category. From a preventive perspective this is not terribly helpful, since clinicians cannot prioritise half their caseload. The practical question is therefore who should the clinicians focus on in order to prevent as many of the most serious incidents as possible? Unless the tool can be refined to be predictive of, for example, the more serious acts of violence, there is a danger that its predictiveness will be statistically impressive but with limited utility from a preventive perspective.

The HCR-20 (Douglas et al 1999) is a 20-item checklist to assess the risk for future violent behaviour in criminal and psychiatric populations. Items were chosen based on a comprehensive review of the literature and input from experienced forensic clinicians. The HCR-20 includes variables which capture relevant past, present, and future considerations and use of the tool is described as an important first step in the risk assessment process. A manual provides information about how and when to conduct violence risk assessments, research on which the basic risk factors are based, and key questions to address when making judgements about risk. Violence is defined as "actual, attempted, or threatened harm to a person or persons". The professional who completes the HCR-20 Coding Sheet must first determine the presence or absence of each of the 20 risk factors according to three levels of certainty (i.e. Absent, Possibly Present, Definitely Present). In some settings, responsibility for the assessment may be divided among several different professionals. The 20 items are divided into three sections: 10 Historical Items (previous violence, age at first violent offence, family and vocational background, etc.), 5 Clinical Items (current symptomatology and psychosocial adjustment), 5 Risk Management Items (release and treatment plan, necessary services and support).

The HCR-20 has generated a substantial research literature and is widely used, at least in forensic contexts. For example, Douglas et al (1999) used both the Psychopathy Checklist and the HCR-20 and found that HCR-20 subscales were moderately predictive of violence in community follow-up of service users over an average period of about two years. Its main disadvantages for routine use are its lack of coverage of risk to self and the breadth of its categories, many practitioners preferring more specific questions.

RAMAS, the Risk Assessment, Management and Audit System is perhaps better tailored to routine use in mental health services (O'Rourke (1995), O'Rourke, Hammond and Davies (1997)). The original RAMAS comprised a 66-item checklist of indicators derived from the clinical and research literature. This checklist has subsequently been extended to 83 items (O'Rourke and Hammond 2000). The RAMAS factors are more specific than the HCR-20, which is advantageous. RAMAS has also been used to generate an interesting psychometric model which allocates the items to one of three scales: the Dangerousness scale, the Mental Instability scale and the Self Harm scale. These scales have been shown to have reasonable internal consistency and there is evidence that the RAMAS scales relate to both subjective clinical judgement and the presenting problem of the service user. However, as the authors admit, the correlations are somewhat lower than might have been hoped: the correlation between clinical judgement and dangerousness score was 0.51, mental instability score 0.24 and self-harm score 0.28. This highlights a practical dilemma: if a scoring system is used then a practical problem is going to arise whenever the score conflicts with the clinician's judgement. Clinicians are potentially placed in an invidious position: if they act on their judgement then they could be accused of ignoring the score generated by a validated measure; whilst the alternative of overriding their judgement and acting on the basis of the score is likely to be even less palatable. There is a sense in which judgements of risk can be said to comprise two distinct judgements. The first judgement concerns whether or not there is a significant risk. The second judgement is concerned with a more precise understanding of the nature of that risk and of exactly what actions should or should not follow. It is plausible that clinicians will find score-based, normative tools more acceptable in the latter context, where they are attempting to make a complex decision and where any additional information is helpful. One response might be to say that in case of conflict further consideration should be given and special care taken over the correct course of action. However, it remains to be demonstrated that this conflict between idiographic and normative approaches can be satisfactorily resolved when idiographic thinking is so fundamental to clinical practice. One practical effect of this conflict is likely to be reluctance on the part of practitioners to take up score-based risk assessment systems, at least as a first-line tool.

In sum, all the tools reviewed have disadvantages when considered as potential first-line risk assessment tools for routine mental health practice. Paradoxically, it is often those features that in more specialist contexts would appear to be strengths (comprehensiveness, depth, complexity, scoring algorithms) that make them poor candidates for routine use. Furthermore, none of the tools include idiographic information. It is perfectly possible to record idiographic data in a standardised way. For example, risk factors such as 'recurrence of circumstances previously associated with risk behaviour' or 'occurrence of early warning signs of relapse' are not included, presumably because they would refer to different actual circumstances or relapse signatures for different individuals. However, from a practical viewpoint such factors are highly relevant to clinical judgements but they are unlikely to be picked out by the standard research strategy of sifting through the literature for relevant predictors.

There is therefore a need for tools which can sit more easily with routine practice: that can facilitate and support risk assessment and risk management but which take better account of the individual's circumstances and history and are better integrated with naturally occurring clinical processes.

2   Developing a Tool: The Infometric Approach

There are two fundamentally different strategies to the development of tools for routine use.

The first strategy is to use the standard development methodology used in research contexts but with a view to ensuring that the resulting tools are economical to use and generic in content. Recognising that tools developed for research purposes do not lend themselves to routine use, there has been a move in the direction of briefer, broader and easy to administer tools, the view being that if tools are brief enough and generic enough then they will be usable in routine practice. This approach has been referred to as 'clinometric' (Wright & Feinstein, 1992). 'It recognises that 'today's new opportunities to measure health status routinely and on a large scale demand the best compromise between traditionally defined psychometric elegance and the new standard of feasibility and practicality' (Stewart et al., 1992). Two well-known examples of such tools in general use are the SF-36 and HoNOS (Wing, Curtis and Beevor 1996). In the Risk context the HCR-20 or RAMAS are examples of this tradition.

Whilst the modified research-based development strategy can result in relatively brief generic tools, there is little evidence that this paradigm leads naturally to tools that gain widespread acceptance as part of routine practice. This is not surprising if one considers the implicit context assumed by research tools. In essence, research tools pre-suppose a context where:

- The data is collected regardless of practical exigencies such as the hustle and bustle of day-to-day practice.

- The assessors are fully trained in use of the tool.

- The assessors are highly motivated to complete the tool in full.

- There is a long interval between data collection and any results/feedback from the data.

- It is not necessary to monitor day-to-day fluctuations in state.

- The tool is not to be used to support day-to-day practical decisions.

- The tool is not to be used as a means of sharing information… etc.

Tools that are the product of these assumptions may be reasonably rigorous from a particular theoretical perspective but are frequently experienced as rigid and impractical. To take a simple example of a practical difficulty, most research-based tools do not provide scope for free text recording of the individual details of risk factors identified for clinical purposes. Hence the practitioner is immediately faced with having to have two parallel recording systems in order to summarise the results of a single assessment. Nor do research tools include scope for the recording of risk management plans. Hence a further document is required for this purpose. So before you know it the risk assessment tool that was to be at the heart of routine practice has become an appendage to the 'real' clinical documentation. The practical benefit of a core document that could be a day-to-day focus of multi-professional practice and communication is lost.

However, the problems with the 'research' approach go beyond mere 'practical' difficulties in their use. The underlying assumptions can also lead to psychometrically inappropriate design. For example, from a research perspective if Variable A is highly predictive of risk and Variable B is equally predictive of risk but Variable A and Variable B are highly correlated then there is usually no statistical advantage in including Variable B in a tool. Hence tools tend to be abbreviated to the minimum possible set of data items. However, in practice this is true only if near-100% data collection of Variable A is guaranteed; and/or if measurement of Variable A is very accurate. If neither of these assumptions can be taken for granted it would be more psychometrically sensitive to build in a degree of redundancy and collect both Variable A and Variable B.[2]

Assessment tools are shaped, whether consciously or unconsciously, by the context in which they evolve. If the aim is to design a tool for a particular context, such as routine use in secondary mental health services, the optimal approach is likely to be to design a tool based upon an analysis of the characteristics of that context, rather than hope that methods that were developed for a quite different context can simply be modified to suit the new one. The effect of adopting an inappropriate methodology is likely to be to arrive at a tool that looks like it might be usable but in fact struggles to find acceptance. There is an important distinction between tools that aren't too much of a nuisance to use routinely if you have to, and tools that are an integrated part of routine practice. If risk assessment is truly to be a cornerstone of good practice, what is required is the latter.

2.1　The Infometric Approach

The alternative approach is to begin with an analysis of the context of routine practice, the processes that occur and the data that is in any event collected and on this basis design a tool which both fits the clinical context and which also has sufficiently robust psychometric properties. This approach we refer to as the 'infometric approach' since it aims to help routinely-collected data become information with measurement properties. The 'infometric' approach combines the principles underlying the development of information tools with those underlying the development of psychometric tools, hence the hybrid term 'infometric'. Such an approach must of course take advantage of what the research literature tells us are important characteristics to look out for, but this is merely one component of the design, and even here research findings need to be filtered through the prism of everyday practice. More broadly, the key features of the infometric approach are that it:

- Follows the naturally-occurring clinical process.

- Integrates recording and measurement.

- Integrates with the service context.

- Provides information feedback.

Based upon application of the above to risk assessment, it is possible to develop a requirements analysis for the design of a risk assessment tool.

---

[2] This is a separate point to the more obvious one that there may be good clinical reasons for collecting both A and B. For example, ability to wash independently and ability to dress independently are highly correlated but this does mean that an OT or nurse only assesses one and not the other.

2.2 Requirements analysis

The key requirements identified are as follows:

1) First, any tool needs to be brief and easy to complete by practitioners with a wide range of skills and experience.

2) The content of the tool should reflect the breadth of factors relevant to risk to self and to others[3]. The literature reviewed focuses on the 'sharp end' of risk, that is homicides and suicides, as these are the areas of greatest concern from both clinical and research perspectives. Nevertheless, the factors that need to be covered by risk assessment in routine practice are much broader. The concept of 'risk to others' includes not just risk of homicide but risk of violence to others (whether homicidal in intent or otherwise), risk of physical abuse of others and even risk of emotional abuse. Similarly, the concept of risk to self includes not just risk of suicide but risk relating to other forms of deliberate self-harm, for example cutting and self-mutilation, as well as risk arising out of reckless behaviour or severe self-neglect. The tool therefore needs to cover all key risk factors. Such risk factors may be either historical or reflected in the person's current state.

3) Risk assessment begins as part of the standard process of clinical and social assessment. In many cases it is evident that there is little or no risk; whilst in other cases risk is a major feature of the presentation. A tool needs to cope with both eventualities. Additionally, the contents of the risk assessment should complement other assessment tools being used.

4) Where it is apparent that there is a potential risk, the severity and nature of the risk needs to be clearly identified and communicated to others involved with the person's care. Central to this is the clinical judgement of risk. Clinical practice is based upon the application of clinical judgement deriving both from the mental health evidence base and the professional experience of the practitioners themselves. Clinical judgement is pivotal to clinical practice. This centrality of clinical judgement needs to be taken on board both in relation to the content of a tool (e.g. pitching the judgements made at a similar level to the judgements made in practice, rather than making them too global) and in the tool's psychometric design (building in the capability to examine the relationship between risk factors and clinical judgement).[4] The judgement of risk should be based upon a considered judgement that takes into account the risk factors and warning signs identified. However, in line with the concept of supporting existing processes rather than inventing new ones, the tool should *not d*etermine how the practitioner interacts with the service user in determining the presence or absence of risk factors, this being left to the professional training of the practitioner.

5) Recording and measurement need to be integrated. The clinical recording of information is usually done in case notes using handwritten free text and has the benefit of capturing the richness and individuality of the person and their specific circumstances.

---

[3] The literature in general focuses on one or the other. A Medline search on risk assessment and psychiatry from the years 1995 to 2002 did not find a single paper (out of 200) that reviewed both risk to self and risk to others. From a scientific point of view this brings with it the danger that the relationship between risk to self and risk to others is not being adequately explored.

[4] The infometric approach thus relies on clinical judgement in a way that, for example, a research instrument may not, since the latter may, for example, include a standardised interview format that could be used to collect data by anyone with suitable training in the instrument. Such questions are explored in Part 2.

Measurement function has features which are typically lacking in clinical recording: a measurement tool selects key features; it describes them in a succinct and standardised way that can be understood by all with training in the use of the tool; it describes those features in a manner that is easy to communicate to others and which enables comparison between one person and another. Finally, measurement tools lend themselves to easy computerisation – a major advantage as information technology moves towards the development of computerised clinical information systems. Thus what is required is a single tool which is both usefully descriptive of the person and capable of fulfilling measurement functions.

6) Where it is apparent that there is a potential risk, plans or interventions need to be put in place to address them. These need to be recorded and easily accessible.

7) The resource implications of any plans or interventions need to be balanced against the resource availability for the service setting in which the risk is presenting. This implies the need for a normative component in tool design – enabling comparison of an individual with other service users.

8) The demands of multi-professional and multi-agency working require that the information collected during assessment be summarised in a form that can be easily communicated to others involved and easily accessed as circumstances require.

9) The tool should potentially support the development of team- or agency-wide risk management strategies as well as individual risk management.

10) The tool should include a measure of 'risk outcome' so that when suitably updated it is possible to explore the characteristics of service users with differing risk outcomes.

3   The FACE Risk Profile

This section describes the main features of FACE Risk Profile. The tool is designed to supplement use of either local assessment/CPA documentation or the FACE Triage, Core or Health and Social Assessment for those service users for whom significant risk to self or others is suspected. Initial description of the tool is followed by commentary on a number of key aspects, indicating key considerations that underpin the design.

3.1   Description of the tool

The Risk Profile occupies a single sheet of paper (two double-sided sheets of A4, folded down the middle). The information recorded divides into the following categories:

Front sheet

This records:

- Administrative data: name of service user, ID no, contact details of care co-ordinator, etc.

- Known previous history of significant risk behaviour?

- Involvement in serious incident or 'near miss' in past three months.

- Current risk status. Overall judgements about the service user's current risk status, including: Risk of violence/harm to others, Risk of suicide, Risk of deliberate self-harm, Risk of severe self-neglect/accidental self-harm, and Risk to child.

- Proneness to relapse.

- Checklist of persons potentially at risk (staff, family member, general public etc.).

Page 2

Page 2 comprises a detailed Warning Signs & Risk Factors Checklist. Factors are grouped under five headings: Clinical symptoms indicative of risk, Behaviour indicative of risk, Treatment and care-related indicators, Forensic history, Personal circumstances indicative of risk. Separate columns of boxes ask the user to indicate whether there is a Risk History related to each item or whether the item is present as a Current Warning Sign. In order to avoid mindless ticking of boxes, the user is able to answer Yes or No at the heading level to these questions. Only if the answer is 'yes', e.g. 'Yes' there is a history of behaviour indicative of risk, does the user complete the relevant section in detail. The time frame for presence of Current Warning Signs is within the past month.

Page 3

Descriptive Account. This includes:

- Checklist identifying person(s) at risk. The user is asked to tick as appropriate a list including: self, partner/spouse, parent, child, staff member, general public, general public, group, other.

- Space for free text description of Current Warning Signs and Risk History.

Page 4

Relapse and Risk Management Plan.

This final part of the Risk Profile is designed to summarise steps to be taken in the event of relapse or signs of relapse; and/or to minimise risk where risk has been identified or suspected. The Relapse and Risk Management Plan contains four sections:

- Steps to be taken if service user fails to attend or meet other commitments. A checklist of possible actions is provided and space provided for free text details.

- Target signs, symptoms, and behaviour suggestive of possible risk/relapse. This space is used to record key behavioural, symptomatic or circumstantial changes that give cause for raised concern about relapse or risk. For example, if a service user has become violent when a relationship has ended; or relapsed when the key worker was away; or reported certain symptoms prior to relapse, these are recorded here.

- Action to be taken in the event of relapse/risk.

  This section is used to record precise details of what should be done in the event of relapse or perceived likely occurrence of risk, including details of the action(s), how immediate action should be, rationale for the plan, who should be notified, etc.

- Buffers against Risk. This final section records actions, people and circumstances that in the past have been associated with successful containment and management of risk, including both specific actions, e.g. contained with increased medication and general factors, e.g. 'seems happiest when in regular contact with brother'.

3.2 Content

Generation of the checklist of risk factors and current warning signs took account of reviews of the literature, policy and other guidance produced by national bodies and factors identified by other instruments. The views of a range of practitioners were sought in seeking to minimise the length of the checklist, the aim being to maximise practical utility without omission of key factors.

Within the Risk Profile only those factors that are most directly indicative of risk are included. More detailed assessment of areas such as severity of symptomatology, physical well-being and disability or interpersonal functioning are assumed to be undertaken in context of the person's broader assessment, with only key indicators recorded in the Risk Profile.

In some instances items have been included that have not been used in previous tools. For example, one important contributory factor consistently identified in enquiries has been poor documentation of risk by professionals involved in care and poor communication of the identified risk to others involved with that service user (e.g. Ritchie et al 1994). A frequent charge made by relatives of those involved in incidents has also been that warnings by relatives or others close to the service user were ignored by clinical staff. For this reason, the factor 'concern expressed by others' was included.

An important feature of the checklist is the inclusion of idiographic factors, albeit expressed in standard terms. Two such factors are: 'Early warning signs of relapse', and 'Recurrence of circumstances associated with risk behaviour'. These are trying to tap into the triggers of risk behaviour in a manner that is more sensitive than ticking a list of objective factors.

In other instances, items were excluded either due to infrequency of occurrence or because they were felt to be difficult to assess routinely, especially at initial assessment (e.g. items with a strong psychological component such as 'denial of problems', and 'unrealistic expectations'; low IQ, excluded partly for lack of acceptability and partly due not being routinely assessed and 'organic' items such as head injury and learning disability).

3.3 The risk status rating scale

The development of a rating scale for the recording of judgements of risk was the most sensitive aspect of the development process. Many clinicians are uncomfortable with the concept of rating risk since they feel it may imply an ability to predict or a depth of knowledge of the service user that may not be present. An initial draft rating scale of 'low, medium or high' risk was therefore rejected. The disadvantage of the 'low, medium, high' approach was similarly reported by the developers of RAMAS themselves, in that they found that it was not completed by one third of respondents. They reported that staff did not feel competent to make a judgement of risk without input from other colleagues and that they were concerned that an imprecise view was registered on a service user form (Hammond & O' Rourke 1997).

In view of the initial problems the risk scale was replaced by a version of the tool that did not explicitly ask for the recording of such judgements. This too, however, was regarded as inadequate and self-defeating: what was the point of a Risk Profile that did not record judgements of current risk? As a consequence, an alternative approach to production of a rating scale was attempted, which proved far more acceptable to clinicians. This was to anchor the points of the rating scale against practical decisions required in the course of clinical practice, e.g. a clinician has to decide whether immediate action is required in order to reduce risk. Based upon this principle, each level of risk rating is linked to a course of action appropriate to that level:

0=No apparent risk. No history or warning signs indicative of risk.

1=Low apparent risk. No current behaviour indicative of risk but person's history and/or warning signs indicate the possible presence of risk. Necessary level of screening/vigilance covered by standard care plan, i.e. no special risk prevention measures or risk management plan are required.

2=Significant risk. The person's history and condition indicate the presence of risk and this is considered to be a significant issue at present, i.e. a risk management plan is to be drawn up as part of their care plan.

3=Serious apparent risk. Circumstances are such that a risk management plan should be/has been drawn up and implemented.

4=Serious and imminent risk. Person's history and condition indicate the presence of risk and this is considered imminent, e.g. evidence of preparatory acts. In these circumstances highest clinical priority should be given to risk prevention.

In making the ratings of risk, clinicians may therefore use the same thought processes as those that go into the clinical decision as to what course of action is required.

A person may be given a positive risk rating in more than one risk category. In making the rating, users are instructed to take into account the following in making judgements of current status: severity of behaviour, regardless of intent; intention; level of impulsiveness in the service user's behaviour; frequency and duration of the behaviour in question; history, especially recent but also more distant behaviour; imminence of the behaviour. Inter-rater agreement on judgements of risk status appears to be high, at least in settings where service users are well-known to staff.

3.4   Use of the tool

The Risk Profile is designed for routine use in mental health services, rather than as a research instrument or a formalised assessment procedure. It is designed to support and supplement clinical practice, not modify it. As the training manual states 'the evaluation, prediction and management of risk will almost certainly require consideration of other information collected as part of the more general clinical and social assessment, and possibly through the completion of specialist assessments (e.g. of personality, attitudes and beliefs, self-esteem, self-control and impulsivity, intellectual function and organic impairment, interpersonal and social circumstances, etc.)'.

The Risk Profile is not designed as a structured interview and does not dictate the way in which the assessment is carried out, this being left to the practitioner's discretion. Although the items covered may serve as a useful 'aide-mémoire' in relation to types of risk factor, the Profile is intended only to be used to record information collected through reasonable and practical enquiry (for example, inspection of any available letter of referral or case notes, the completion of a sensitive clinical interview with the service user and significant others, etc.).

3.5   Order of completion

The basic approach to completion of the Risk Profile is as follows:

1) Determine whether the person has a history of significant risk behaviour or whether there are any obvious current indications of risk.

2) If the answer is 'No' to both questions in 1) then complete page 1 only.

3) If the answer is 'Yes' to either of the questions specified in 1) then complete Page 2 (the checklist of risk factors and warning signs) and Page 3 (the associated descriptive account).

4) Having completed pages 2 and 3, complete the judgements of risk on page 1.

5) Complete the Risk management plan on page 4.

There are a couple of important points here. First, the tool is designed so it can be used universally across mental health services. However, there is no point in insisting that

detailed risk assessment be undertaken with literally every service user. Assessment should be proportionate to need. The procedure above enables a certain level of risk assessment to be completed and documented without forcing practitioners through an excessively lengthy procedure.

Secondly, the basic aim of the procedure is to ensure that practitioners have taken full account of past and present risk factors when making their judgements of risk. Hence, pages 2 and 3 should be completed *before* the judgements of risk. At the same time, however, the front page is designed as an at-a-glance summary of the findings of the risk assessment. It is extremely useful to be able to see a summary of risk status, persons at risk, previous involvement in serious incidents, etc. without having to browse through a detailed document. Additionally, the front sheet may be used as a data sheet for electronic recording (many services send off a copy of the front sheet only for entry onto a central risk register and keep the full document in the casenotes for reference). Thus, whilst at first sight it may seem slightly confusing that the front page is not completed first, there are sound practical reasons for the design.

3.6 Training procedure

Training tools are provided with the tool. These include detailed instructions relating to completion of each section of the Profile, standardised vignettes of sample service users with completed 'gold scores' for both checklist items and risk status items, and sets of 'prompts' that may be used to elicit specific items of information in the clinical context. A typical training session lasts about half a day and includes introduction to the tool and approach, completion of a standardised vignette followed by discussion and feedback. Trainers may be trained in a day.

Training also covers broader issues relating to use of the tool. For example, there is often anxiety about the making of judgements of risk. The user is instructed that judgements should be made on the basis of such reasonable assessment and enquiry as it has been possible to complete. This statement recognises that clinical, ethical and practical requirements are such that it is often either impractical or inappropriate to gain a conclusive answer to all items before making the clinical judgement as to whether there is a perceived risk. The assessor is only expected to have made such enquiry as would appear to be appropriate to the clinical context: this is a clinical judgement of comparable complexity to many other clinical judgements made routinely. Thus if someone presents with a relatively straightforward phobia or social problem it will obviously be inappropriate to enquire at great length over risk history – a few screening questions would usually suffice. Similarly, it may be impossible to gain all the relevant information from someone in an acutely distressed or florid state.

3.7 Involving service users

One of the broader benefits that might be hoped for from introduction of a standard risk recording procedure is enhanced communication, at least between practitioners and ideally also between practitioners and service users. In an interesting qualitative study, Edwards et al (1998) concluded that 'the standardization of the language of risk was felt to have potential benefit in making professionals consistent in their appreciation of risks and communication with each other. Between professionals and service users, standardized language was thought inappropriate or insufficient because of contextual variation in communication and interpretation of risk information by service users.'

In developing the Risk Profile a decision had to be made as to whether to attempt to develop a tool whose language was as transparent to service users as to practitioners. In the event, this was attempted to a limited degree only, for two reasons. First, time and resources made establishing the necessary consultation process difficult. Secondly, and most importantly, practitioner anxiety concerning risk assessment is so high that it seemed a risky undertaking to attempt to modify language to enhance acceptability in such a sensitive area. What was done instead was to try to minimise the use of professional language without compromising clinical requirements.

Having done this there was some concern that users would find the explicit use of such terms as 'Risk of physical harm to others', or 'Command hallucinations' off-putting. In fact, however, local consultation with user groups received such favourable feedback that service user representatives became involved in both local and national training in use of the tool. Rather than being put off by explicit language, users seem positive about 'calling a spade a spade' and the tool's holistic inclusion of personal circumstances and the service user's and carer's perspective has also been praised.

3.8 Data considerations

The Risk Profile can generate data which:

(i)   Enables description of the patterns of naturally-occurring risk factors and levels of risk in a clinical population.

(ii) Enables exploration of the characteristics of judgements of risk and the relationship between the presence of risk factors and those judgements.

(iii) Enables exploration of issues relating prediction and judgements of risk.

(iv) Enables monitoring of changes in risk status.

Thus, as well as supporting day-to-day clinical practice with individuals, the tool is a potentially rich data source which can be used both to explore key issues relating to risk such as prediction and the nature of clinical judgements; and to inform broader risk management strategies; and also to inform further development of the tool itself.

Use of the data to support such exploration forms the next part of this paper.

4   The Data

This section presents some initial data collected using the Risk Profile. The majority of the data were collected in the context of a national audit project based at the British Psychological Society's Centre for Outcomes Research and Effectiveness at University College, London and funded by the Department of Health's National Centre for Health Outcomes Development (Clifford 2001). This involved collection of data on a sample of service users in both in-patient and community mental health settings across 25 UK mental health services.

4.1   Dataset

The measures used included descriptive variables and other tools described below:

Registration and Background Dataset

This included basic details such as age, gender, housing and economic status as well as brief details of history of contact with mental health services, details of current episode and chronicity of illness.

The Health and Social Assessment

This is designed to provide a comprehensive assessment of Health and Social functioning in the following domains:

Psychological well-being	Physical well-being
Activities of daily living	Interpersonal relationships
Social circumstances	Family and informal carers
Risk	Response to care

The Health and Social Assessment includes two scaled global judgements of risk, referred to as 'Risk to self' and 'Risk to others' which use the same scale as the Risk Profile. This data is included in some of the analysis.

How Are You?

This brief questionnaire asks the service user to report upon their quality of life and mental and physical well-being (Clifford et al 2005). It covers all FACE assessment domains and has been validated against the practitioner-based FACE assessment tools.

4.2   Site selection

The sites were self-selecting in that choice of participation was dependent upon the sites themselves volunteering following a publicity mailshot. Sixteen sites supplied the data presented below.

4.3   Data collection requirements

All sites were asked to collect the Registration and Background data and Risk Profile data on a minimum of 50 service users. The Risk Profile was repeated in some cases but this was not mandatory.

Sites were permitted to collect data on:

(i)   New service users.

(ii) Existing service users coming up for review.

The Background dataset permitted distinction between 'new' service users and others already on the practitioners' caseloads.

Use of How Are You? was optional.

4.4   Demographic characteristics of the population

The sample reported upon comprises 497 service users included in the National Outcomes Audit of Severe Mental Illness.

Risk Profile data was actually collected on nearly 800 service users but the smaller sample is focused upon in the analysis as this subset also had complete data on other relevant datasets, such as the Health and Social Assessment and How Are You?

There were 218 women and 279 men in the sample. The sample had an average age of 33.4yrs (s.d. 23yrs).

Four hundred and forty-four service users were of White-British or White-Irish origin (89.3%), 16 ((3.2%) were Black-Caribbean or Black-African and 12 (2.4%) were of Asian origin. Five percent were rated as 'Other'.

About 25% (124) of the population lived alone, 22% lived with their partner or spouse, 13.5% lived with their parents and 6.0% lived with their children only. The majority of the remainder (28.4%) were either in long-stay hospital or supported accommodation.

Fifty-three percent of the sample were single, 21% married or cohabiting and 22% divorced, separated or widowed.

4.5   Clinical and service characteristics

The service users were receiving care in a range of in-patient and community settings as shown in Table 1 below. The sample typically had chronic mental health difficulties, with nearly three fifths of the population having had mental health difficulties for five years or more and less than 15% having had difficulties for less than one year.

Table 1: Location of risk-assessed service users		
Service setting	**N=**	**%**
In-patient	43	8.7
Intensive care ward/Secure unit	12	0.8
Long-stay ward	33	6.6
Rehabilitation ward	61	12.3
Day hospital	73	14.7
Out-patient department/CMHC	38	7.6
Supported accommodation	30	6.0
Home	146	29.4
GP Premises	54	10.9
Other	7	1.4

4.6 Diagnosis

Table 2 shows the diagnostic breakdown on the sample. By far the most common diagnosis was Schizophrenia (35.6%) followed by Depression (18.7%).

Table 2: Primary diagnosis	
Schizophrenia	177 (35.6%
Depression	82 (18.7%)
Schizoaffective disorders	22 (4.4%)
Neuroses/anxiety disorders	18 (3.6%)
Bipolar disorder/mania	17 (3.4%)
Acute psychosis	11 (2.2%)
Other mood disorder	8 (1.6%)
Schizophrenia with substance misuse	7 (1.4%)
Other schizophrenic type disorder	5 (1.0%)
Alcohol and drug misuse	2 (0.5%)
No diagnosis	15 (3.0%)
Non-psychiatric disorder	4 (0.85)
Not reported/not known/not assessed	129 (26%)

4.7   Reference group

All service users were assigned to one of two 'reference groups' for purposes of analysis and benchmarking. The two reference groups are referred to as 'SMI' (severe mental illness) and CMH (common mental health problems). This process used an algorithm that takes into account both diagnosis and clinical and social characteristics of service users. The algorithm has been shown to have high agreement with practitioner assignment to reference group; and to result in far higher rates of allocation than use of diagnosis alone (diagnosis is often not recorded). The distinction between SMI and CMH reference groups has been shown to be fundamental to the analysis of clinical data (Clifford 2003). It differentiates service users on a very wide range of health and social variables; outcome measures; and both style and level of response to self-report measures (ibid.).

Three hundred and ten service users belonged to the SMI reference group and 187 belonged to the CMH reference group. Within the SMI reference group the most common diagnosis was Schizophrenia (n=177). Within the CMH group the most common diagnosis was Depression (n=78).

5 Psychometric Properties of the Risk Profile

This section presents some basic psychometric data on the Risk Profile.

5.1 Internal consistency of risk domains

Analysis was undertaken of the internal consistency of the items classified within risk sub-domains used to classify risk factors and current warning signs in the Risk Profile. The results are shown in Table 3 below, which shows the Cronbach's alphas for the items as classified under the headings within the questionnaire.

Table 3: Internal consistency of Risk Profile domains	
N=497	**Cronbach's alpha**
Clinical risk indicators	0.81 (items=16)
Behavioural risk indicators	0.81 (items=22)
Treatment-related risk indicators	0.81 (items=16)
Forensic risk indicators	0.74 (items=8)
Personal circumstances indicative of risk	0.81 (items=10)

All alphas are satisfactory, confirming that the classification is psychometrically satisfactory as well as conceptually reasonable.

An alternative approach to grouping the factors was also explored, classifying the items according to whether they related primarily to risk to self or risk to others. In a number of cases (e.g. concern expressed by others) this was not possible and the item was included within both groups. Classification in this way also permitted comparison with another established tool in this area, the RAMAS (O'Rourke, 1995). The results, shown in Table 4, show that classification in this way results in higher Cronbach alphas of around 0.9, these alphas being slightly higher than the comparable alphas for the RAMAS.

Table 4: Internal consistency of Risk Profile domains classified by self and others		
N=497	**Cronbach's alpha**	**RAMAS**
Items relating to harm to others (all)	0.92 (items=28)	0.85 (items=29)
Items relating to harm to others (historical)	0.90 (items=14)	n/a
Items relating to harm to others (clinical and behavioural only)	0.86 (items=12)	n/a
Items relating to self-harm (all)	0.90 (items=16)	0.81 (items=12)
Items relating to self-harm (historical)	0.90 (items=8)	n/a
Items relating to self-harm (clinical and behavioural only)	0.84 (items=10)	n/a

The raised alphas are at least in part a function of the greater number of items and thus do not necessarily reflect an enhanced method of classification. Nevertheless, the results suggest that classification in this way may be helpful in, for example, generating separate sub-scores for risk to self and risk to others. However, at a practical level further consideration would need to be given to the issue of overlap of risk factors and the related question of the extent to which risks to self and to others can be clearly separated. This question is further explored in Chapter 7.

5.2 Validity of the Risk Profile

The validity of the Risk Profile was assessed in a variety of ways, described below.

Face validity

The Profile was derived from both a review of the literature and wide consultation with representatives of all mental health professions. This, coupled to the fact that clinicians were willing to use the instrument to summarise a considerable number of assessments are suggestive of reasonable face and content validity.

Criterion validity

Criterion validity was examined during the process of constructing sample case descriptions for training purposes. Professionals who had completed the Risk Profile were interviewed about their judgements of risk status and asked to explain what aspect of the service user they were considering in making a judgement of, say, serious risk to others. These answers were recorded and presented back (blind) to a group of five professionals. Out of 10 descriptions generated the group agreed in 92% of cases that the description related to those items nominated by the original assessor.

Construct validity

Construct validity was examined by comparing sub-groups of service users differentiated on diagnostic or service criteria which would be expected to predict different patterns of risk scores. For example, it would be expected that service users suffering from depressive disorders would score more highly on items relating to self-harm than service users suffering from anxiety disorders. Using this procedure a number of simple hypotheses were confirmed ($p < 0.01$):

- In-service users scored more highly on all items than out-service users.

- In samples of in-service users and out-service users matched for severity of psychotic symptoms, in-service users scored more highly on problems related to self-harm or harm to others.

- Service users with a diagnosis of schizophrenia scored more highly on items related to psychotic symptoms.

- Service users with a diagnosis of depression scored more highly on items related to depressed mood, poor self-esteem, self-harm and psychosomatic symptoms.

- Service users on higher tiers of the Care Programme Approach scored more highly on items relating to self-harm and harm to others.

Concurrent validity

The Risk Profile scores were correlated with domain scores and items scores on the FACE Health and Social Assessment (Table 5). All correlations shown were statistically significant at the 0.01 level.

Table 5: Correlation between Risk Profile domain scores and Health and Social assessment axes						
N=497	**Psycho-logical**	**Behav-iour**	**Mental health**	**Inter-personal**	**ADLs**	**Index**
Risk status	0.54	0.65	0.40	0.34	0.35	0.30
Risk factors total	0.37	0.38	0.27	0.39	0.29	0.39
Symptoms	0.39	0.36	0.31	0.33	ns	0.33
Behaviour	0.41	0.46	0.27	0.41	0.37	0.45
Forensic involvement	ns	Ns	ns	ns	ns	ns
Personal circs	0.29	0.28	0.24	0.28	0.18	0.28
Risk to self now	0.42	0.54	0.26	0.24	0.22	0.33
Risk to others now	0.49	0.56	0.34	0.34	0.26	0.44

As would be expected, the correlations were highest in relation to the Psychological domain, which includes items on current severity of behaviour. If the Psychological axis is broken down into separate behaviour and symptom sub-scores, then risk status correlates about 0.7 with the behaviour sub-score and only around 0.4 with the 'mental health' sub-score which focuses on severity of symptoms. The question of the nature of judgements of risk is explored more thoroughly in the following chapter.

All other Risk Profile domains correlated with all other Health and Social Assessment domains in the range 0.3-0.4 except for Forensic involvement which did not correlate significantly with any of the Health and Social assessment domains.

5.3   Overall level of risk

The question whether it is meaningful or useful to talk about an overall risk score is an important one, since it has potential practical implications for both clinical and resource management. This issue was explored initially by examining the distribution of risk scores across the participating sites.

The Risk mean score on the combined 'Risk to self' and 'Risk to others' items for the whole population was 0.64 (s.d. 0.72, s.e. 0.02). However, as can be seen from Figure 1 there was substantial variation in levels of perceived risk across the participating sites, with mean scores per site ranging from around 0.2 to nearly 1.5. Of the 15 sites, the four sites with the lowest mean scores fell outside the 95% confidence limits of the six sites with the highest scores, indicating a high level of variability.

This variation was explored further by grouping sites according to the type of care setting in which the assessed person was receiving care. The results are shown in Figure 2 below. It can be seen that the three 'community'-based locations of care GP (n=51), Home (n=150) and Day Hospital (n=71) have barely overlapping confidence limits for levels of risk and all fall outside the confidence limits of the two in-patient settings (acute in-patient ward n=64, intensive care ward n=13). The two in-patient settings have similar mean levels of overall risk with far greater variation in the intensive care ward, perhaps reflecting the small n. The mean risk score in the G.P. setting was 0.19 (n=64) compared to the mean risk score in secondary care of 0.62 (n=433).

Figure 1: Overall Level of Risk by Site
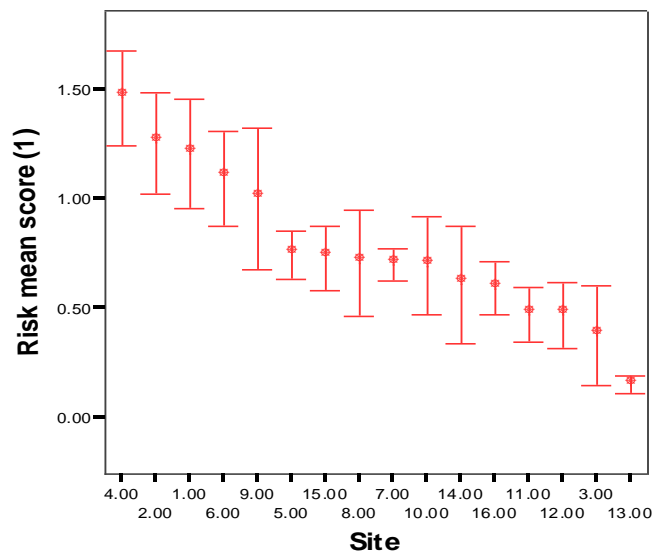
Figure 2: Level of Risk in different locations of care
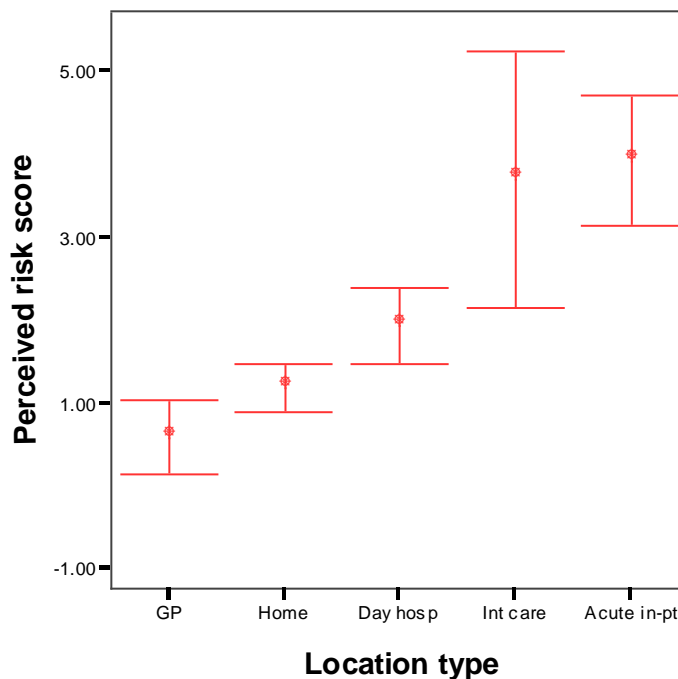
These results, which operate at a very global level, are nevertheless encouraging both in further demonstrating the capacity of the tool to distinguish between different settings and in suggesting in a preliminary way that the concept of a global risk score may have some utility.

5.4 Inter-rater reliability of judgements of risk to self or others

Two reliability studies were undertaken, one at a day hospital (n=25) and one in an acute ward (n=20), using the ratings of risk to self and risk to others. The inter-rater agreement relating to judgements of risk to self and risk to others as measured by weighted kappa was about 0.9 in both studies.

5.5 Summary

The results presented in this chapter provided encouraging preliminary data on the psychometric properties of the Risk Profile. The classification of risk factors within the Profile is internally consistent and that the Risk Profile sub-domains compare favourably with those of another established tool, the RAMAS (O'Rourke, 1995). Construct validity tests confirmed that the tool distinguishes along expected lines between clinical populations, whether classified by diagnosis or location of care. Concurrent validity tests also demonstrated modest but consistent correlations of around 0.4 between numbers of risk factors identified and severity scores on Health and Social Assessment domains and a higher correlation of 0.65 between perceived risk status and scoring on behaviour items on the Health and Social Assessment. Although full-scale inter-rater reliability trials were not conducted on the whole tool, the inter-rater reliability of the key judgements of risk was good.

Subsequent chapters provide further data germane to psychometric assessment of the tool, including further analysis of the tools discriminatory capability (Chapter 6) and analysis of the structure of judgements (Chapter 7) which speaks to the internal consistency of the tool as used in practice.

6 Profiling Risk

This chapter presents basic descriptive data on risk factors and levels of risk for the population assessed. As well as providing basic illustration of the data generated by the tool, such data also point the way towards some of the more complex analyses undertaken in later chapters.

6.1 Specific risk factors

First, it is straightforward to examine specific risk factors identified on the Risk Profile. The mean number of factors identified was 8.5 (s.d. 9.47). The mean number of historical factors identified was 5.7 (s.d. 6.6) and the mean number of current risk factors identified was 2.8 (s.d. 3.8).

However, as would be expected there was a large difference between those service users perceived to present a risk and those not perceived to present a risk. The mean number of risk factors for those who were not judged to present a risk (i.e. had a perceived risk score of <2 on all risk judgement items) was 5.6, with a mean historical risk factors score of 4.2 and a mean current risk factors score of 1.4. For those were judged to present a risk the mean risk factors score was 17.5, with a mean historical risk factors score of 10.5 and a mean current risk factors score of 6.9.

Overall, the likelihood of a risk factor occurring previously was at least twice that of it being a current risk factor.

On a separate item, nearly two thirds of the population (319) were rated as having a history of significant risk behaviour.

As would be expected there was a large variation in the incidence of different risk factors. At one extreme, factors such as self-neglect, prodromal signs of relapse and concern expressed by others were present in over 20% of cases. At the other extreme, items such as morbid jealousy and admission to Special Hospital were present in only around 1% of cases. As would be expected, the more specific and more grave the risk factor the less frequent its occurrence. The full table of frequencies is shown in Table 6 below.

Domain/item	Current		History	
Clinical symptoms indicative of risk	**201**	**40.4%**	**285**	**57.3%**
Early signs of relapse	66	13.3%	155	31.2%
Ideas of harming others	39	7.8%	101	20.3%
Ideas of self-harm/suicidal ideation	65	13.1%	136	27.4%
Delusions	101	20.3%	184	37%
Command hallucinations	42	8.5%	89	17.9%
Morbid jealousy	7	1.4%	15	3.0%
Impulsivity/lack of impulse control	86	17.3%	157	31.6%
Other clinical symptoms indicative of risk	28	5.6%	33	6.6%
Behaviour indicative of risk	**213**	**42.9%**	**330**	**66.4%**
Preparation to harm others	16	3.2%	47	9.5%
Suicide attempts	16	3.2%	78	15.7%
Plans to commit suicide	21	4.2%	63	12.7%
Wandering	38	7.6%	80	16.1%
Other deliberate self-harm	18	3.6%	68	13.7%
Threats/intimidation	57	11.5%	131	26.4%
Severe self-neglect	109	21.9%	197	39.6%
Physical harm to others	32	16.4%	121	24.3%
Domestic risk	73	14.7%	118	23.7%
Drug/alcohol abuse	54	10.9%	119	23.9%
Other unsafe/risk-taking behaviour	28	5.6%	77	15.5%
Forensic involvement suggestive of risk	**17**	**3.4%**	**63**	**12.7%**
Conviction for violent or sexual offences	9	1.8%	43	8.7%
Admission to Special Hospital	4	0.8%	13	2.6%
Admission to Secure Unit/Intensive care	14	2.8%	40	8.0%
Other forensic involvement suggestive of risk	1	0.2%	11	2.2%
Personal circumstances indicative of risk	**203**	**40.8%**	**267**	**53.8%**
Recent severe stress	81	16.3%	126	25.4%
Concern expressed by others	124	24.9%	203	40.8%
Recurrence of circumstances associated with risk behaviour	84	16.9%	138	27.8%
Abuse/victimisation by others	40	8.0%	94	18.9%
Social isolation	109	21.9%	161	32.4%
Rootlessness	27	5.4%	37	7.4%

Table 6: Frequency of occurrence of specific risk factors

Treatment-related risk indicators	112	22.5%	217	43.7%
Discontinuation of medication	56	11.3%	166	33.4%
Failure to attend appointments	42	8.5%	132	26.6%
Unplanned disengagement from services	25	5.0%	101	20.3%
Compulsory admission	33	6.6%	139	28.0%
Placement upon supervision or risk register	24	4.8%	34	6.8%
Supervised discharge	9	1.8%	19	3.8%
Restriction order	11	2.2%	14	2.8%
Conditional discharge	6	1.2%	8	1.6%

Finally, Table 7 below shows the percentage of the population with problems in the following areas: risk factors related to treatment, forensic involvement, behaviour indicative of risk to self and others, and symptoms indicative of risk to self and others. It can be seen that risk factors in the forensic category occurred much less frequently than those in other categories. However, of course, such risk factors are of an especially serious character.

Table 7: Frequency of occurrence of risk factors in each domain	
% with risk factors related to treatment (e.g. discontinuation of medication, failure to attend appointments, unplanned disengagement from services)	12.5%
% with forensic involvement (e.g. conviction for violent or sexual offences, special hospital, secure unit, other forensic involvement suggestive of risk)	3.6%
% with behaviour indicative of risk to self (e.g. suicide attempts, other deliberate self-harm, wandering, self-neglect, domestic risk, drug/alcohol abuse, other unsafe behaviour, plan to commit suicide)	26.5%
% with behaviour indicative of risk to others (e.g. physical harm to others, threats/intimidation, preparation to harm others)	26.6%
% with symptoms indicative of risk to self (e.g. ideas of self-harm)	38.3%
% with symptoms indicative of risk to others (e.g. delusions, command hallucinations, ideas of harming others, morbid jealousy, impulsivity)	27.7%

6.2 Perceived risk status

Table 8 below summarises the number of individuals judged by practitioners to be a risk to self or others, grouped according to the various categories of risk judgement included in the Risk Profile. For convenience, the population is banded into three groups: those presenting no or low apparent risk; those presenting with a significant risk and those presenting with a serious or serious and imminent risk. The bottom two rows show summed scores, generated by summing the scores on all risk judgements and all risk judgements relating to risk to self respectively. In these cases the categorisation was derived by banding the summed scores as indicated.

Table 8: Judgements of risk on the Risk Profile			
N=497	**No or low risk**	**Significant risk**	**Serious risk**
Risk to others	92.2%	5%	2.8%
Risk of suicide	93.2%	4.3%	2.5%
Risk of severe self-neglect	83.5%	11.3%	5.2%
Risk of other deliberate self-harm	93.8%	4.3%	1.9%
Risk to self or others*	80.2%	10.2%	9.6%
Risk to self**	83%	12.6%	4.4%

* Summed score of <3=no or low risk, 3 or 4=significant risk, and >3= serious risk.
** Summed score of <2=no or low risk, 2 or 3=significant risk, and >3= serious risk.

At least 80% of service users were perceived to be low or no risk in each category. By far the most frequently perceived risk was that of severe self-neglect, occurring in 16% of cases.

Risk to others was present in about 8% of cases and risk of suicide about 7% of cases.

In total, about 8% of service users were perceived to be a risk to others and about 7% perceived to be at risk of suicide. Taking into account all risk judgements, about 20% of individuals were perceived to present a significant or greater risk to self or others.

6.3 Overall level of risk to others

Levels of risk, as measured by judgements of risk to self or others, were analysed according to site and locations of care. Chapter 5 demonstrated that overall score differentiated between care settings. Further analysis explored the separate contribution of perceived Risk to self and Risk to others.

First, risk to others scores were examined. The distribution of levels of risk to others by site is shown in Figure 3 below. These show a similar pattern of distribution to the global scores examined in the previous chapter.

Figure 3: Level of Risk to Others by site
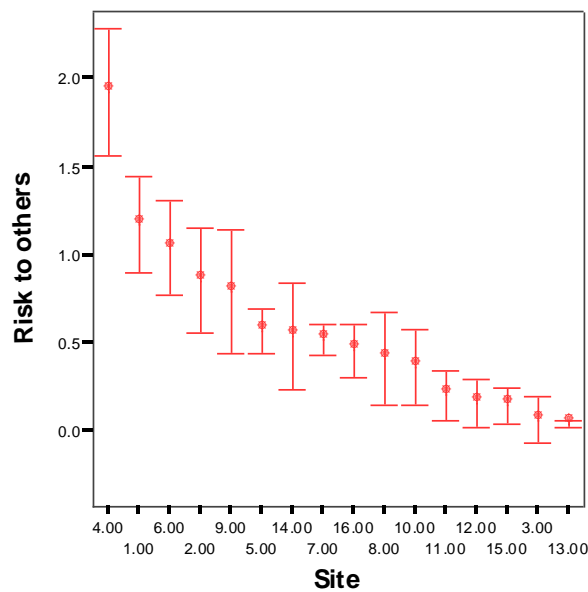
In fact, only one site changed its ranking by more than three places. This suggests that risk to others is highly determinate of average risk score.

Analysis of levels of risk to others in different care settings was also undertaken and the results shown in Figure 4 below. The locations again differed from each other along expected lines with the two in-patient settings scoring more highly. Interestingly, there was relatively little difference between the Day hospital and Acute in-patient scores but a large difference in mean score between the intensive care and acute in-patient scores, to such an extent that even though the confidence limits of the intensive care ward were far broader than the acute in-patient ward, the two did not overlap.

Figure 4: Perceived Risk to Others in different locations of care
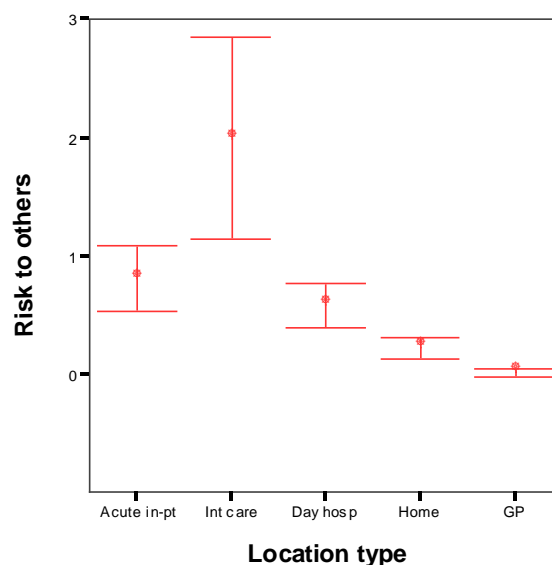
6.4 Overall level of risk to self

A similar set of analyses examined comparable patterns in relation to Risk to self. Figure 5 below shows the mean levels of risk to self across sites. There was again some variation in scores between and within services. However, the pattern of variation is substantially reduced when compared to either the overall risk score or risk to others, with only the highest and lowest scoring sites falling outside the confidence limits of the remainder. This suggests that level of risk to self is a less important determinant of either overall risk or allocation or presentation to care setting than risk to others. This was confirmed by examination of perceived risk to self by locations of care (Figure 6).

Figure 5: Level of Risk to Self by site
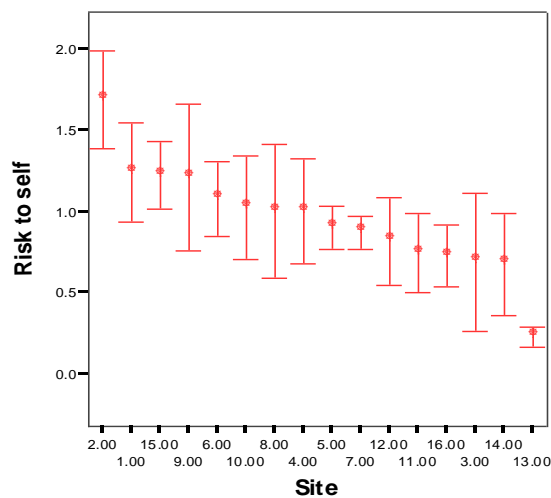
Figure 6: Perceived Risk to Self in different Locations of Care
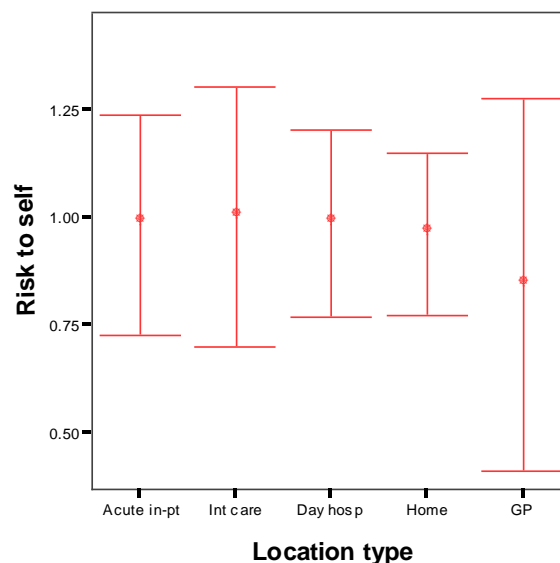
It can be seen that there is very little variation in overall level of risk to self across locations of care, with the greatest variability being in the GP setting and even the mean scores for acute in-service users being similar to those service users being seen at home.

6.5 Levels of risk by reference group

A further analysis looked at proportions of service users presenting with different levels of risk to self and others in the SMI and CMH populations.

Table 9 shows that while the majority of service users posed no or a low risk to others, many more service users in the SMI population than in the CMH population belonged to the high risk categories (16% compared to 3.4%). Thus the findings in relation to Risk to others confirm the broad pattern identified when examining overall pattern of risk in the previous chapter. The tool thus appears to discriminate well between both settings and reference groups.

Table 9: Levels of risk to others by reference group		
Level of risk	SMI (n=310)	CMH (n=187)
No apparent risk	58.1%	84.6%
Low apparent risk	26.9%	12.2%
Significant risk (needed plan)	9.5%	1.6%
Serious risk (plan implemented)	4.7%	1.0%
Serious and imminent risk (highest clinical priority given to risk prevention)	0.8%	0.6%

When the levels of risk to self were broken down by reference group (Table 10), nearly 20% of the SMI group fell into the high risk categories compared to about 13% of the CMH group, a difference but considerably less marked than in relation to risk to others. More service users posed a risk to themselves than to others in both populations.

Table 10: Levels of risk to self by reference group		
Level of risk	SMI (n=310)	CMH (n=187)
No apparent risk	38.0%	60.1%
Low apparent risk	42.3%	27.2%
Significant risk	12.5%	9.5%
Serious risk	5.9%	1.8%
Serious and imminent risk	1.3%	1.4%

6.6 Discussion

The data illustrate how a tool such as the Risk Profile can begin to provide an understanding of a clinical population in a service context.

The tool demonstrated wide variation in average levels of risk across service settings along expected lines and proved highly sensitive to differences in perceived levels of risk even within related settings such as in-service users, day hospital and intensive care wards. It was further demonstrated that this variation arose substantially from perceived differences in risk to others rather than risk to self. This is an interesting finding, especially in the light of the national policy commitment to reduce suicide rates, since it shows that there is a high correlation between perceived risk to others and level of service provision but apparently minimal correlation between perceived risk to self and level of service provision.

When the data was partitioned by reference group, as expected there was a higher incidence of risk to others amongst the SMI reference group. This also applied to risk to self, with one in five of the SMI population being considered a significant risk to self as against one in eight of the CMH population. In fact, levels of risk to self were higher than risk to others in the SMI population, a feature perhaps under-recognised by policy emphasis on avoidance of risk to others.

The data also provide an interesting preliminary insight into the multi-variate nature of perceptions of risk and the complexity of the judgement facing practitioners. Those judged to present a significant or greater risk averaged about 17 risk factors compared to the average number of risk factors observed in the population as a whole of about eight.

This scale of task is further illustrated by the finding that about one in five individuals were perceived to present a significant risk to self or others. It would seem unlikely that resources would permit special measures to be taken with 20% of the clinical population.

Further exploration of the nature of clinical judgements of risk forms the subject of the next Chapter.

7   Judgements of Risk

The complex nature of judgements of risk tends to produce two contradictory responses: on the one hand that 'it is not possible to accurately predict risk' and on the other a desire for risk assessment tools to do just that. The problem with both of these responses is that neither is empirically-based: there is little research exploring the nature of judgements of risk in mental health or how they are arrived at. The present study therefore presented an unusual opportunity to examine the relationship between several sets of variables: practitioner judgements of risk, practitioner-rated presence or absence of specific risk factors (both historically and recently), practitioner judgements of severity, the occurrence or otherwise of serious incidents involving harm to self or others; and the service user's self-report on their thoughts and feelings as provided by the 'How Are You?' questionnaire. Enhanced understanding of such relationships could potentially cast a valuable light on practical questions concerning the validity of practitioner-based and service user-based risk assessment and approaches to the management of risk. This and the next chapter explore such issues. This chapter focuses on the practitioner-based data and examines the relationship between the recorded occurrence of risk factors, practitioner judgements and the occurrence of serious incidents. The following chapter then explores the relationship between practitioner judgements and service users' self-report.

7.1   Domain scores and practitioner judgements

The Risk Profile includes factors relating to both risk to self and risk to others as well as judgements of current risk status. This structure permits exploration of the relationships between objectively-occurring risk factors and subjective judgements of levels of perceived risk.

As a preliminary analysis two variables were calculated: a total risk factors score, created by summing the number of risk factors and current warning signs checked in completion of the Profile; and a total perceived risk score, generated by summing the scores on all four practitioner judgements of risk. The correlation between these two variables was 0.64 (n=672, p<0.001). This appears to suggest that judgements of risk have at least a partial quantitative basis. However, if only those perceived to be a significant or greater risk were considered, the correlation dropped to 0.31, compared to a correlation of 0.61 for those not perceived to be a risk. Thus although there was a general relationship between numbers of risk factors and perceived level of risk, *this was lower for the population of most concern*, namely those at risk.

The next step was to examine relationships between scores on different sub-sections of the questionnaire and more specific judgements of risk. Scores were generated for each sub-section of the risk factors and warning signs checklists by summing the number of positive answers in each column. Thus, for example, the 'Behaviour indicative of risk' score was generated by summing the number of Behaviours indicated as present now with the number of behaviours that had occurred historically. In effect, therefore, each item within each heading received a score of 0, 1, or 2 depending on whether it was absent, occurred historically, present at the time of assessment, or both. The scores so generated were then correlated with three of the clinician judgements of levels of current risk: the score on the 'Risk to others' item; the score on the 'Risk of suicide' item; and a

total 'Risk to self' score generated by summing the scores on the 'Risk of suicide', 'Risk of other deliberate self-harm' and 'Risk of severe self-neglect' items.

The full results are shown in Table 11 below. As might have been expected, scores on Behaviour correlated most highly with individual clinical judgements of risk status, being at just under 0.6 for both the 'Risk to others' score and the Risk to self overall score.

Table 11: Correlations between Risk Factor Scores and Judgements of Risk			
N=497	Risk to others judgement score	Risk of suicide judgement score	Risk to self judgements score
Clinical symptoms indicative of risk score	0.48**	0.42**	0.55**
Behaviour indicative of risk score	0.58**	0.33**	0.58**
Treatment-related indicators score	0.26**	0.17**	0.32**
Forensic involvement suggestive of risk score	0.34**	0.015	0.049
Personal circumstances indicative of risk score	0.27**	0.41**	0.52**

Interestingly, the correlation with the 'Suicide risk' judgement item was substantially lower at 0.3, although still significant. However, in this case the correlation with clinical symptoms and personal circumstances was higher at 0.4. This perhaps suggests that clinicians are placing greater weight on suicidal ideation, other clinical symptoms (such as hopelessness) and personal circumstances in assessing suicide risk. Of note also was the correlation of 0.52 between 'Personal circumstances indicative of risk' and the overall 'Risk to self' judgements score.

In the next set of analyses three scores were used: the total risk factors score, generated by summing the total number of items in the checklist receiving a positive response; a total 'Risk to self' score, generated by summing scores on the items within the 'Clinical symptoms' and Behaviour' sub-sections which had obvious direct relevance to 'Risk to self' (e.g. suicidal ideation); and a total 'Risk to others' score, generated by summing scores on the items within the 'Clinical symptoms' and Behaviour' sub-sections which had obvious direct relevance to 'Risk to others' (e.g. Physical harm to others). These correlations are shown in Table 12.

Both the Risk to others and Risk to self total scores correlated more highly with the Risk to others judgement and Suicide risk judgement respectively than any of the domain scores reported in the previous analysis, suggesting that sheer quantity of risk factors is closely related to practitioner judgement. Interestingly, the Risk to others total score had a correlation of nearly 0.5 with the Risk to self judgement score. This relationship between risk to self and risk to others is explored further in section 7.6.

For purposes of comparability the corresponding RAMAS correlations are shown in the table. The Risk Profile compared favourably, producing a correlation of 0.61 with the Risk to others judgement score (compared to RAMAS 0.51) and a correlation of 0.66 with the Risk to self judgement score (compared to RAMAS 0.28).

Table 12: Correlations between sub-groups of risk factors and judgements of risk			
N=497	Correlation with Risk to others judgement score	Correlation with Risk of suicide judgement score	Correlation with Risk to self judgements total score
Risk factors total score	0.49**	0.39**	0.58**
Risk to self total score	0.41**	0.50**	0.66**
Risk to others total score	0.61**	0.29**	0.48**
RAMAS[5] Risk to others risk factors score	0.51	-	-
RAMAS Risk to self risk factors score	-	-	0.28

7.2 Correlation of past and present factors with judgements of risk

The judgement of risk takes account of both the past and the present in making a judgement about the future. It is of interest therefore to examine the respective correlations for previous occurrence of a type of factor and current occurrences of that type of factor. Separate scores were therefore calculated for past and present factors and correlations with judgements of risk produced. These are shown in Table 13 below.

---

[5] The RAMAS validation included only a single clinical judgement of 'risk'. The correlations reported here are with separate judgements relating to risk to self and risk to others. It is possible that the RAMAS correlations would have been higher had these been separated out.

Table 13: Correlation of current factors and historic factors with judgements of current risk			
N=497	Correlation with Risk to others judgement score	Correlation with Risk of suicide judgement score	Correlation with Risk to self judgements score
Risk to self factors total score	0.41**	0.50**	0.66**
Risk to self now total score	0.41**	0.59**	0.72**
Risk to self history total score	0.35**	0.37**	0.52**
Risk to others factors total score	0.61**	0.29**	0.48**
Risk to others now total score	0.60**	0.33**	0.53**
Risk to others history total score	0.50**	0.21**	0.37**

The Risk to self now score correlated more highly (0.72) with the judgement of current risk to self than the Risk to self history score (0.52). The Risk to others now score also correlated more highly with the judgement of current risk to others (0.6) than the Risk to others history score, though the difference was smaller (0.6 against 0.5).

A final set of correlational analyses undertaken combined the three classificatory axes previously explored separately: risk factors relating to risk to self and relating to risk to others; risk factors relating to behaviour and those relating to mental health; and currently occurring factors versus historically-occurring factors. In this set of analyses, eight scores were created covering all possible combination of these three factors. The results are shown in Table 14 below. The highest correlation was between the Risk to self now behaviour score and perceived Risk to self (0.67). This compared with a correlation of 0.58 between the Risk to others now behaviour score and perceived Risk to others.

It was expected that current behaviours would correlate more highly with clinical judgements of risk than current symptoms. However, there was little difference in correlation between current behaviour scores and current mental health scores in relation to either risk to self or risk to others. In both cases the correlations were around 0.6.

Table 14: Correlations between behaviour and symptoms scores and judgements of risk to self or others (n=497)				
Practitioner judgement scores **Factor scores**	**Correlation with total perceived risk score**	**Correlation with Risk to others judgement score**	**Correlation with Risk of suicide judgement score**	**Correlation with Risk to self judgements total score**
Total risk factors score	0.61**	0.44**	0.43**	0.56**
Risk to self now score (mental health)	0.52**	0.35**	0.57**	0.61**
Risk to self now score (behaviour)	0.67**	0.38**	0.51**	0.67**
Risk to self history score (mental health)	0.42**	0.31**	0.37**	0.47**
Risk to self history score (behaviour)	0.49**	0.33*	0.20**	0.46**
Risk to others now score (mental health)	0.55**	0.51**	0.35**	0.52**
Risk to others now score (behaviour)	0.57**	0.58**	0.23**	0.41**
Risk to others history score (mental health)	0.40**	0.40**	0.24**	0.38**
Risk to others history score (behaviour)	0.40**	0.53**	0.13**	0.28**

These analyses suggest that current state is used by practitioners as the primary indicator of risk. Whilst this might be no surprise, it has to be weighed against the general finding of behavioural science that past behaviour is the best predictor of future behaviour. Again, however, the issue of timeframe is crucial: past behaviour might in general terms be a better predictor of future behaviour; but current state might be a better predictor of behaviour in the next few days or few weeks.

7.3 Modelling clinician judgements

In order to explore further the relationship between clinical judgements and the occurrence of risk factors, three regression models were generated. The first regressed all individual items in the Risk checklists on the clinician judgement score for Risk to self and Risk to others. The second regressed the domain scores only on the clinician

judgement score. The third tested the hypothesis that breaking the mental health and behaviour domain scores into separate Risk to self and Risk to other components would enhance the model.

The results of these analyses are shown in Table 15. Multiple r values are high, with a multiple r of 0.84 between the total practitioner risk status score and the item scores. This provides further support to the evidence presented in Chapter 5 that initial consideration of risk factors leads to a consistent approach to succeeding judgements of current risk status.

Table 15: Regression of item-based scores on to practitioner judgements of risk			
N=497	**Item-based regression model**	**Domain-based regression model**	**Enhanced domain-based regression model[6]**
'Total Risk' clinician judgement	0.84 (r-squared=0.70)	0.66 (r-squared=0.43)	0.77 (r-squared=0.59)
'Risk to others' clinician judgement	0.81 (r-squared=0.66)	0.56 (r-squared=0.31)	0.70 (r-squared=0.49)
'Risk of suicide' score and clinician judgement	0.80 (r-squared=0.63)	0.45 (r-squared=0.20)	0.66 (r-squared=0.44)
'Risk to self score' clinician judgement	0.80 (r-squared=0.62)	0.62 (r-squared=0.38)	0.75 (r-squared=0.56)

Domain-based multiple r values were somewhat lower than the item-based values. A drop would be expected, but it is notable, for example, that the multiple r for risk of suicide drops substantially from 0.80 to 0.45. The multiple r values for the enhanced model (which re-classifies risk factors according to whether they are relevant to risk to self or risk to others) raised values although not to the levels of the item-based model. Thus it is not so much the grouping into domains as such but the method of grouping that leads to loss of predictiveness. Thus, as in previous analyses, the regression values provide further evidence of a strong, quantitative effect of number of risk factors influencing judgements of risk to self and others whilst at the same time indicating that individual risk factors rather than sheer number of factors also play an important part.

7.4 Previous serious incidents and judgements of risk upon initial assessment

Having examined practitioner judgements of risk in relation to identified risk factors, it was then possible to examine the relationship between judgements of risk and the prior occurrence of risk behaviours, as measured by the items on the occurrence of serious incidents in the Health and Social Assessment. The assessment contains two items, one asking whether the person has been involved in a serious incident involving harm to self in the three months prior to assessment, and another asking a similar question

---

[6]This column was generated by splitting symptoms and behaviour into self and others scores for each of current and history.

regarding involvement in incidents involving harm to others. For ease of analysis the answers to the two questions were collapsed into a single yes or no answer.

Table 16 below shows the risk status of service users who were involved in a serious incident involving harm to self or others in the three months prior to initial assessment.

Table 16: Perceived levels of risk and serious incidents at initial assessment			
	No or low risk	Significant risk	Serious risk
No incident in past three months	282	38	22
Serious incident in past three months	38	7	13

If an individual was involved in a serious incident in the previous three months, the likelihood of their being assigned to the significant or serious risk categories was about one in three. If an individual was not involved in a serious incident in the previous three months, the likelihood of their being assigned to the significant or serious risk categories was about one in six. Thus the presence of a recent serious incident resulted in a doubling of the likelihood of being placed in the significant or serious categories.

Taking the serious category alone, about 6% of service users with no recent incident were placed in the serious category, compared to 22% with an incident. Thus the presence of a serious incident resulted in about a fourfold increase in likelihood of being placed in the serious risk category.

The likelihood of allocation did not vary significantly when those with incidents relating to harm to self and harm to others were considered separately.

The further question of the relationship between judgements of risk and the subsequent occurrence of serious incidents is examined in Chapter 8, section 8.4.

7.5 Judgements of risk and judgements of severity

A further area of exploration is the relationship between judgements of risk and judgements of severity. In many circumstances it is tempting to take judgements of severity as a proxy for judgements of risk, on the grounds that severity of illness is likely to correlate with level of risk. The relevant analyses are shown below. First, Table 17 shows the correlation between Risk now score and overall severity score on the FACE Health and Social Assessment. This is 0.5. The correlation with the behaviour scale within the Health and Social Assessment is 0.6 and with the Symptoms scale around 0.4. However, if the same correlations are examined for those service users who are judged to be a significant risk then the correlations fall away. The behaviour score still has a correlation of 0.46 but the overall severity score has a correlation of 0.27.

Table 17: Correlations between judgements of risk and judgements of severity			
	Overall severity score	**Behaviour severity score**	**Symptoms severity score**
Risk now score (all)	0.504**	0.605**	0.411**
Risk now score (all judged to be a significant risk)	0.273**	0.459**	0.288**

When Risk to self and Risk to others are separated (Table 18), correlations are about the same for all service users. However, the correlations for those service users judged to be a risk fall substantially, so much so that the Risk to self now score has a non-significant correlation of 0.16 with the overall severity score and the Risk to others now score has a correlation of about 0.3 with the overall severity score. The correlation with the Behaviour severity score remains at about 0.42 for risk to others but falls to 0.2 for risk to self. These findings are not in themselves surprising since judgements of severity take into account current state and behaviour rather than previous behaviour. However, they do make clear that risk and severity are two quite different concepts and that no easy algorithm is likely to generate proxy scores for clinical judgements of risk. They also illustrate the clinical complexity of the situation: faced with two service users, one of whom is relatively well now but has a serious risk history and another who is more severely unwell now but who has a lesser risk history how is the practitioner to determine who is most at risk?

Table 18: Correlations between judgements of risk and judgements of severity (risk to self)			
	Overall severity score	**Behaviour severity score**	**Symptoms severity score**
Risk to self now score (all)	0.462**	0.518**	0.343**
Risk to self now score (all judged to be a significant risk)	0.163	0.206*	0.186**
Correlations between judgements of risk and judgements of severity (risk to others)			
	Overall severity score	**Behaviour severity score**	**Symptoms severity score**
Risk to others score (all)	0.372**	0.467**	0.230**
Risk to others score (all judged to be a significant risk)	0.319**	0.425*	0.232**

7.6 Relationship between judgements of risk to others and of risk to self

A final set of exploratory analyses in this area was to look at the relationship between judgements of risk to self and judgements of risk to others. This produced some surprising and perplexing findings. The relationship between practitioner judgements of suicide risk and judgements of risk to others was different for different sub-populations

(see Table 19). Within the SMI group as a whole there was no correlation, but within the CMH group there was a correlation of 0.4. However, within the group judged to present a risk (mostly SMI) the correlation between practitioner judgement of risk of suicide and judgement of risk to others was a remarkable −0.79 (p<0.0001).

Table 19: Correlations between practitioner judgement of suicide risk and judgement of risk to others	
Population	**Correlation**
All (n=331)	0.22 (p<0.0001)
All SMI (n=211)	0.054 (p=0.44)
All CMH (n=104)	0.40 (p<0.001)
All judged to be a risk (n=32)	-0.79 (p<0.0001)
All judged not to be a risk (n=298)	0.15 (p=0.009)

Thus whilst in general terms an increase in perceived suicide risk is moderately correlated with an increase in perceived risk of harm to others, within the group who are actually perceived to pose a significant risk to themselves or others judgments of the two types of risk are very strongly *negatively* correlated. On this basis, it would appear that practitioners believe, or at least act as if they believe, that past a certain threshold as risk to others increases risk of suicide decreases and vice versa.

7.7 Discussion

The purpose of the analyses described was to gain a better understanding of clinical judgements of risk. Given this objective, some caution must be exercised in interpretation of the data, which are based upon correlations between the recorded occurrence of risk factors and clinical judgements. The finding of a correlation does not necessarily imply *causation* in the sense that one cannot infer that correlations point to actual clinical thought processes. The finding of a high correlation between a group of risk factors and a clinical judgement does not mean that clinicians use their awareness of those factors in making their judgements. For example, the high correlation between 'Personal circumstances indicative of risk' and judgements of risk does not mean that clinicians take such factors into account when making their judgements of risk – although it might be reasonable to infer cautiously that they do.

Having said that, the correlational findings consistently demonstrate that the tool generates scores that correlate highly with clinical judgements of current risk. It should be recalled that the checklists do not include any graded rating of severity, frequency or intensity of the factors listed – so that achieving correlations of 0.6-0.7 is of note.

Although there was a strong relationship between numbers of risk factors and perceived levels of risk, enhanced correlations were achievable through regression modelling, suggesting that clinicians assign different weightings to different items and/or different domains in making their judgements. This view was also confirmed by the finding that of the types of risk factor recorded, behaviour rather than symptoms tended to be more predictive of practitioner judgements. Similarly, scores on current rather than past risk factors correlated more highly with practitioner judgements, as would be expected.

The data indicate that generating separate 'risk to self' and 'risk to others' scores based upon risk factors drawn from all domains is the best method of 'predicting' clinician judgement. Scoring systems based on this method could easily be built into electronic implementations of the tool.

More generally, the data provide strong validation for the completion procedure described in 3.5, whereby the practitioner makes judgements of risk based upon having completed the checklists and descriptive account of risk factors and warning signs. This appears to result in a consistent approach to judgement – a finding confirmed further by the increased likelihood of higher risk ratings of those involved in a prior serious incident. It is of interest that only one in three such service users were placed in this category – it would not have been surprising if the number had been twice as high. This confirms that practitioners' judgements of risk referred to the relatively immediate future (since the proportion would be expected to be much higher if practitioners were predicting, for example, whether the service user would *ever* be involved in a further serious incident).

A further important finding was the relatively low correlation between severity scores and judgements of risk, especially amongst those considered to be at risk. The results suggest that whilst severity scores pertaining to current behaviour may be taken as at best a modest proxy for risk (correlating around 0.5), global severity ratings have only weak correlations. Thus, global scores generated from severity-oriented tools which are themselves global in nature, such as, e.g. HoNOS or GAF, should not be taken as proxies for risk.

Against this background of apparent consistency, an important finding was that although there was a general relationship between numbers of risk factors and perceived level of risk, this was considerably lower for the population of most immediate concern, namely those perceived to present a risk. There are a number of possible explanations for this. It may be that introducing more sensitive scaling in the rating of risk factors would help. However, this would lengthen the process of completion and close inspection of the Risk Profile does not lend much support to this view. For example, take an item such as 'Ideas of harming others'. In theory, this item could capture anything from occasional angry thoughts of punching someone on the nose to continuous preoccupation with murderous plans. However, the more severe end of this spectrum is also likely to be captured elsewhere within the Risk Profile by a whole range of items such as Impulsivity, Physical harm to others, Preparation to harm others, Threats/intimidation, Conviction for violent offences, History of compulsory admission, etc. Thus, whilst a more carefully graded scale could enhance sensitivity, there are reasons to believe that the enhancement would not lead to capture of much additional information and would therefore be insufficient to explain the large drop in correlation.

A second and more plausible explanation is that from a clinical viewpoint there is an important distinction between identifying the presence of a significant risk and deciding how much of a risk someone actually presents at a particular moment in time. These are quite different problems. It would be quite understandable for there to be a high level of agreement between practitioners on who constitutes a possible risk but lower levels of agreement as to who presents the most risk. If so, this would mitigate against consideration of 'risk' as a simple linear variable.

A further complicating factor is that of time. Identifying someone as a risk is one thing, identifying when they are likely to engage in risk behaviour is another. An analogy here might be with hills and volcanoes. Identifying which of the Earth's bumps are likely to be at risk of eruption is fairly straightforward; predicting when and whether a particular volcano is going to erupt within a certain timescale is extremely difficult and can only be done with any degree of accuracy by close inspection of current geology, ongoing monitoring of activity – and even then prediction is extremely difficult and not at all accurate. This analogy also clarifies the relationship between past and present factors in predicting risk. Past behaviour may well be the best guide to the future (in general terms); but to get any real predictive handle on the *immediate* future you need to focus on present behaviour.

The question of the timeframe or 'sell by date' for judgements of risk is a difficult one from a methodological point of view. It is tempting to specify a time frame (e.g. a month, next three months) for judgements in order to clarify the nature of the judgement that the practitioner is being asked to make. However, this may well produce more problems than it solves, since there are many individuals for whom it is impossible to say whether they are a specific risk or not in the next month but who in general terms are a definite risk and need a risk management plan. If such individuals (e.g. service users with a history of violence towards staff) are rated as 'no current risk' this will be highly misleading. The Risk Profile attempts to address such issues in the definition of the rating scale, which distinguishes between there being a risk that requires a risk management plan (minimum rating 2=significant risk), there being a risk in general terms but no current specific indication (low apparent risk); and there being a current risk of sufficient specificity that something has to be done to mitigate it at present (minimum rating=3).

The complexity of assessing level of risk amongst those deemed to be a risk is further illustrated by the disconcerting finding that the correlation between practitioner judgements of risk to self and risk to others varies considerably according to clinical sub-group. Although the cell size is small (and one would prefer to have the results replicated to feel absolutely confident) this would appear to be a highly specific phenomenon, perhaps confined to the SMI group. It is important to observe that it is not confined to judgements of risk to others since about 40% of the 'at risk' group were judged to be at risk of suicide. The most plausible explanation is that having identified a major risk, practitioners focus on that to the exclusion of other possible risks. If so, there is a danger that they systematically underestimate these other risks.[7] If this is an accurate interpretation, it strongly suggests the need for corrective training which could result in enhanced risk detection and management.

Given the complexities identified in assignment of levels of risk amongst those perceived to be a risk it is perhaps not surprising that actuarial methods have struggled to provide accurate prediction. What then are the outcomes of people judged to be a risk? We shall turn to this in the next chapter.

---

[7] This issue is further explored in the next chapter.

8   Risk Outcomes

The data permit examination of two types of 'risk outcome'. The first is change in risk status, i.e. was the perception that levels of risk dropped over time in the assessed population? Secondly, a more direct outcome measure is the subsequent involvement of service users in serious incidents. The former provides a more clinically-oriented perspective on the natural ebb and flow of risk in relation to clinical status, whereas the latter provides the type of measure more beloved of those interested in prediction.

8.1   Change in perceived levels of risk

Although the full Risk Profile was only completed once in many cases, global ratings of risk status were completed both initially and at follow-up, thus permitting examination of change in risk status based upon comparison of mean risk status scores. The results are shown in Table 20 below.

Table 20: Change in Risk item scores T1 to T2								
Domain/ item (n=551)	T1 mean score	T2 mean score	% change	p =	n with initial problem	T1 mean score	% change	p =
Risk	0.69	0.56	20%	0.000	-	-	-	-
Risk to self	0.49	0.42	14%	0.000	300	1.52	35%	0.000
Risk to others	0.87	0.70	19%	0.059	176	1.46	32%	0.000

In the population as a whole, percentage change was 14% and 19% on the Risk to self and Risk to others items respectively. However, only change on the Risk to self item was statistically significant, in part because of a lower incidence of initial positive ratings on the Risk to others item and in part because of higher scores on the part of some service users upon repeat assessment. Percentage change over time was broadly similar in both reference groups.

The table above summarises the global findings for the population as a whole. However, the people of most interest are those initially judged to be a risk. After all, it is those individuals for whom the service is likely to implement some kind of risk management procedure. When only those service users with an initial positive score on the Risk to others item are considered, the percentage change rises to 32% and the change easily reaches statistical significance. Overall, observed change in those with an initial positive score was 34% (p=0.0001).

8.2   Individuals who changed

Population mean scores do not necessarily provide the best way of viewing change in risk status. Consideration of how many *individuals* changed from being perceived as presenting a risk to not presenting a risk or vice versa is perhaps more meaningful.

Taking the 'significant risk' rating (value=2) as the critical threshold below which an individual was perceived to be not a risk, of the 56 people who were perceived to

present a significant or greater than significant risk to others at Time 1, 27 of these were still perceived to present a significant or greater risk to others at Time 2. However, there were in total 44 people who were perceived to present a significant or greater than significant risk to others at Time 2. Thus about half had moved below the threshold but 17 individuals not previously perceived as presenting a risk to others had moved above the threshold in the same period. In addition, of the 56, half (28) were also perceived to present a significant risk to self at Time 1. This had reduced to 16 by Time 2.

Considering the same data in relation to risk to self, there were 103 people who were perceived to present a significant or greater than significant risk to self at Time 1. Thirty-six of these were still perceived to present a significant or greater risk to self at Time 2. However, there were a total of 68 people who were perceived to present a significant or greater than significant risk to self at Time 2. Thus whilst about two thirds of the original 'at risk' cohort had moved below the threshold about half of these had been replaced by further individuals (n=32) not previously perceived as presenting a risk to themselves. Additionally, of the 103, 28 were also perceived to present a significant risk to others at Time 1. This had reduced to 15 by Time 2.

So, in total 131 individuals (21%) presented a risk to themselves or others at Time 1, of which around 40% (54) were perceived to continue to present a risk to themselves or others at Time 2. Thus the substantial drop in average level of risk in the population to some extent masks the fact that on the one hand many service users considered to be a risk upon initial assessment were nevertheless still considered a risk six months later whilst others who had not previously been considered a risk now were.

8.3 Controlling for overall change in state

Anticipating such considerations, the study asked practitioners to provide a global rating of whether the person had improved, deteriorated or stayed about the same upon completing the follow-up assessment. This enabled the data to be sorted independently of the risk ratings.

The effect of disentangling those who improved from those who deteriorated during the study is shown in Table 21 below, which bands the population according to whether they were judged to be better, worse or about the same upon repeat assessment. Those rated as 'better' had higher risk scores initially and these dropped by about 30%. The scores of those rated as worse began lower but went up by about 17%, reaching a level similar to the starting point of those who improved.

Table 21: Change in perceived risk in populations rated as improving, staying the same and deteriorating					
	Time 1	Time 2	% change	% change (all)	P= (t-test)
'Better' (n=150)	0.89	0.62	30%	19.6%	0.000
'Worse' (n=102)	0.72	0.84	17%	19.6%	0.04
'About the same' (n=167)	0.65	0.50	23%	19.6%	0.001

Interestingly, the scores of those rated as about the same started off the lowest and dropped by a further 23%. This reflected an apparently general tendency for clinicians to be more likely to perceive deterioration than improvement, in the sense that percentage improvement had to be relatively high for a rating of 'better' to be given, whereas a relatively small deterioration tended to correspond with a judgement that the person had got worse.

8.4 Serious incidents and judgements of risk at follow-up

A more concrete measure of outcome than changes in practitioner perception of risk status is whether or not the person had been involved in a serious incident since initial assessment. Table 22 shows the risk status of service users who were involved in a serious incident involving harm to self or others in the three months prior to repeat assessment.

Table 22: Perceived levels of risk and serious incidents at follow-up			
	No or low risk	Significant risk	Serious risk
No incident in three months prior to repeat assessment	293	40	24
Serious incident in three months prior to repeat assessment	27	5	11

If an individual was involved in a serious incident in the three months prior to repeat assessment, the likelihood of their having previously been assigned to the significant or serious risk categories was about two in five (42%). If an individual was not involved in a serious incident in the previous three months, the likelihood of their having been previously assigned to the significant or serious risk categories was less than one in five (18%).

If a service user was considered 'no or low risk' upon initial assessment then the likelihood of their being subsequently involved in an incident was about one in twelve (8%). If they were considered to be a risk upon initial assessment then the likelihood of subsequent involvement in an incident was one in five (20%). Thus practitioner assignment to significant or serious risk categories was linked with a subsequent rate of incidents two and a half times greater than assignment to low risk categories.

If a service user was rated a 'serious risk' upon initial assessment then the likelihood of their being subsequently involved in an incident was a little less than one in three (28%). If they were considered to be a low risk upon initial assessment then the likelihood of subsequent involvement in an incident was about one in eleven (9%). Thus practitioner assignment to the serious risk categories was linked with a rate of incidents about four times greater than assignment to low risk categories.

Thus, practitioner assignment of higher risk was moderately predictive of subsequent involvement in a serious incident. However, despite this, *three quarters of the service users subsequently involved in serious incidents were not rated initially as a serious risk and nearly two thirds were rated as low or no risk*.

8.5 Discussion

It would be expected that levels of risk would vary over time, depending upon the clinical state of the service user, their adherence to treatment and other life circumstances. Perceived high levels of risk would naturally lead to measures designed to lower risk, measures which at least in some cases would be expected to be successful. Similarly, it is no surprise, given the fluctuating nature of mental illness, that some individuals were judged to be more of a risk upon repeat assessment than upon initial assessment.

The Risk Profile proved sensitive to change, with percentage change in those service users considered a significant or greater risk high at 32%. About half those individuals considered to present a risk to others and a third considered to present a risk to self were still perceived to present a risk at follow-up. However, other individuals not previously considered to present a risk were now considered to do so, the replacement rate being in the order of 50% in both cases. In the clinical situation, this 'replacement rate' within the baseline population would also be supplemented by new cases, some of whom would also present a risk. Thus a net reduction in levels of risk over time within an overall clinical caseload may be at best modest and difficult to perceive (or achieve), even if a particular cohort were to be improving.

The Holy Grail of risk assessment has always been prediction. What do the data tell us about prediction? Those judged to be a risk upon initial assessment were about three times more likely to be involved in a subsequent incident than those judged initially not to be a risk. This would seem to indicate a modest degree of predictiveness in initial clinical judgements. However, it is impossible to quantify how accurate the initial judgements were because where a serious risk was indicated, measures would presumably have been taken to minimise occurrence of incidents. Such actions may range from monitoring someone more frequently to physically restraining the individual from harming themselves or others. Whatever is done, it is reasonable to assume that at least a proportion of incidents are prevented in any given time period. However, in the absence of a comparison with a control group for whom no preventive action is taken, which it would be ethically unacceptable to effect, it is impossible to quantify the level of prevention achieved. This is confounded further by variations in availability of local resources – in one retrospective study, Bindman et al (2000) found that key workers believed that in 73% of cases the (at risk) service users had not been prioritised to receive more services than they otherwise would have done. Thus enforced reliance on naturalistic data inevitably leads to an underestimate of the degree of predictiveness of judgements of risk – and an underestimate that would appear to be itself subject to considerable (and unknown) natural variation. Thus, true 'predictiveness' is probably substantially higher than would appear but is intrinsically unknowable.

The limitations of data capture can also present a misleading picture. In the present study caution is required in interpreting the data on 'prediction' due to the study design, which recorded judgements of risk at Time 1 and Time 2 but did not record fluctuations in perceived levels of risk that may have occurred in the interim. It is perfectly possible that an individual rated as 'no risk' initially but subsequently involved in a serious incident might have been rated as 'high risk' in the week prior to the incident (given the opportunity). Without a whole series of studies showing the effect of frequency of data capture the impact of this is again impossible to estimate. This caveat also indicates that

the concept of 'prediction' always needs to be qualified with a statement specifying over what time frame. In the present study the timeframe was determined by the study design, and was thus a six-month period. However, in clinical practice this may well not be suitable and what may be required is a system for ongoing monitoring and updating of risk status.

The caution that needs to be exercised in working with such data is exemplified by the fact that one in five service users regarded as a serious risk were involved in subsequent incidents. This might be interpreted as indicating a limitation in preventive strategies. Perhaps this is just as well, though, since if preventive strategies had been 100% successful the data would have appeared to suggest a negative relationship between initial judgement of risk and subsequent behaviour!

The whole issue of prediction is further complicated by the fact that three quarters of all service users subsequently involved in an incident were not initially rated as a significant risk. Whilst this finding may be misleading due to risk ratings being relatively short-term and the relatively long interval between first and last assessment, it nevertheless highlights a cruel dilemma in deciding upon preventive strategies. On the one hand, the fact that it is possible to identify high risk individuals suggests that it is sensible and ethical to focus on preventing risk in that sub-group; whilst on the other, the fact that most service users subsequently involved in incidents were not initially perceived to be a risk suggests that this can be of limited utility and that perhaps better early detection of risk would be equally or more effective.

Whilst precise interpretation of the data may be debatable, it is reasonable to infer that the prevention of serious incidents is a task unlikely to be achievable to any major degree within current service configuration. Even leaving aside the group who were rated initially as no risk, two fifths of those involved in serious incidents had previously been assigned a high enough risk rating to indicate that preventive action was being taken, suggesting that even enhanced monitoring of those perceived to be at risk has only a limited impact upon the number of serious incidents.

Whilst supporting those who have found that practitioner judgements of risk are of predictiveness comparable to actuarial predictions (Fuller and Cowan (1991)), the data therefore also provide some grounds for caution over the possible accuracy of clinical prediction and preventive strategies. What is most important is that these issues are well-understood by those charged with managing risk. The Risk Profile produces suggestive data concerning what might be called the natural limits of prediction in a routine context. These need to be borne in mind in developing risk management strategies and can only be further elucidated through the introduction of structure to routine data collection.

This vignette of the dynamics of risk in mental health services shows that in conducting risk assessment, practitioners are faced with a multi-layered and complex task with effects that may be difficult to ascertain. Although arguably crude, the type of quantification facilitated by the Risk Profile provides some potential insight into this situation. It also opens up a number of important questions that could be explored using the tool. For example, it would be possible to examine the relative effectiveness of different clinical risk management strategies at the population level by tracking their

impact on replacement rates. This might cast a very different light on how best to manage risk when compared to individual risk management strategies.[8]

To date, 'risk management' seems to have focused exclusively on individual-based methods backed up by cohort-based research. It has to be questioned whether sole reliance on these methods is the best approach. A good example is the approach to outcome measurement used in the present study, which in method is research-oriented rather than management strategy-oriented. The outcome measures presented here are cohort-based, i.e. they track a defined cohort of individuals over time and measure their change in status or involvement in incidents, etc. Whilst cohort outcomes, either of a cohort as a whole or of individual members or sub-groups of the cohort are of course of interest, from a broader management perspective what may be more of interest are what might be called 'service' or 'team' outcomes. The aim of a service-wide risk management strategy might reasonably be to reduce levels of risk in the service as a whole. Whilst this will involve minimising risk in those individuals already identified as a risk, it could also involve such strategies as controlling levels of referral so that replacement rates are <1; ensuring that average levels of risk within particular teams do not go above certain thresholds; ensuring that variation in levels of risk across similar clinical settings do not vary more than a certain amount; ensuring that variation in levels of risk across care co-ordinators or consultant psychiatrists do not vary more than a certain amount, etc. Such 'outcomes' are partly the product of change in levels of risk presented by individuals or cohorts, but are also the product of broader management of levels of risk of newly-presenting service users; management of distribution of risk across the service, etc.

---

[8] For example, it would be possible to examine the extent to which whether resource-intensive individual case management increases replacement rates through diverting resources away from management of broader caseloads.

9 Comparing Service User and Practitioner Perceptions of Risk

So far, we have concentrated on practitioner perceptions of risk. Although such perceptions show a high level of internal consistency and appear to be based upon a consistent response to the occurrence of certain risk factors, important questions remain about the validity of such judgements and their relationship to the perceptions of service users themselves. This chapter explores the relationship between practitioner and service user views of risk.

9.1 Sample

The sample comprised a subset of 331 service users for whom the Risk Profile was completed and who also completed the How Are You? self-report questionnaire (Clifford et al 2002). Of these, 196 were male and 131 female. Two hundred and eleven belonged to the SMI reference group of whom the majority (126) had a diagnosis of schizophrenia. One hundred and twenty service users belonged to the CMH reference group, of whom about a half (59) had a mood disorder.

9.2 Method

How Are You? contains four items that factor analysis undertaken as part of previous studies suggested relate to risk (Clifford et al 2002). Two of these items related to risk to others and two related to risk of suicide:

I have heard voices that frighten or upset me.

I have had thoughts about harming others.

I have made plans to end my life.

I have had thoughts about or felt like harming myself.

Factor analysis on the original version of How Are You? grouped these four items together into a single factor. However, further analysis on a revised version of How Are You? separated the two sets of items, the former two being placed in a 'Risk' factor and the latter two falling under the more general 'How You Have Been Feeling' mental health factor (Clifford et al 2002). Despite the re-classification the latter two items appear to relate directly to the possibility of suicide. For analytical purposes, therefore, the combined scores on the two pairs of items were used to create two variables: HRUSELF which summarised the person's self-report of suicidal ideation and HRUOTH which summarised the person's report of ideas relating to harm to others.

The relationship between practitioner and self-report measures of risk were then explored through four sets of analyses. The first set of analyses examined relationships between practitioner-based and self-report-based scores in the sample population as a whole. The second set of analyses examined the level of agreement between practitioners and service users on the presence of risk. The third set of analyses focused on the relationship between practitioner and service user perceptions solely on those service users whom practitioners had identified as posing a significant risk to others or at risk of suicide. The fourth and final set of analyses examined the relationship between

risk of suicide and risk to others, in relation to both practitioner and self-report items, casting an interesting light on the previous analyses.

9.3 Relationship between practitioner variables and service user-based variables related to risk in the whole sample

The correlations between the scores were examined, as was the correlation between the individual items. Given the high correlations reported in chapter 7 between practitioner judgements of risk and occurrence of risk factors such as suicidal ideation, a reasonably high correlation was expected between the practitioner and self-report scores, even allowing for service users not always reporting thoughts of harm to self or others to practitioners.

The Pearson correlation between the practitioner judgement of Suicide risk and the self-reported risk to self score (HRUSELF) was 0.29 (significant at 0.01 level). Correlations with individual How Are You? items were 0.29 with the item 'I have had thoughts about or felt like harming myself' and 0.23 with the item 'I have made plans to end my life' (both significant at 0.01 level). The correlation between these two self-report items was 0.56. Although not as high as might have been expected, the hypothesis was thus confirmed.

The Pearson correlation between the practitioner judgement of Risk to others and the self-reported risk to others score (HRUOTH) was far lower at 0.13 (significant at 0.05 level). Correlations with individual How Are You? items were 0.037 with the item 'I have heard voices that frighten or upset me' (non-significant) and 0.16 with the item 'I have had thoughts about or felt like harming someone' (significant at 0.05 level). Thus although there was a correlation between self-report and practitioner perception, the correlation was rather small.

The minimal correlation of the item relating to voices with judgement of risk to others suggests that it should probably not have been included in the analysis – its sole impact being to lower the correlation with the practitioner judgement. For this reason How Are You? was scanned for other candidate items of possible relevance, the most obvious one being 'I have felt angry or irritable'. However, the correlation of this item was even lower at 0.063. A correlational analysis was then run on all How Are You? items but no single item had a correlation of >0.1. The further analyses reported upon below therefore only used the score on the single item 'I have had thoughts about or felt like harming someone'.

9.4 Analysis by reference group

A further set of analyses divided the population according to reference group. It is known that the response on self-report questionnaires of service users belonging to the SMI and CMH reference groups differs even when severity and content of presenting problem are controlled for (Clifford 2003). It was therefore plausible that systematic response differences in reference group were resulting in lower correlations or a different pattern of findings than might otherwise have occurred.

When the above analyses were re-run for the SMI reference group, the Pearson correlation between the practitioner judgement of Suicide risk and the self-reported risk to self score (HRUSELF) increased from 0.29 to 0.39 (significant at 0.01 level).

Correlations with individual How Are You? items increased to 0.39 with the item 'I have had thoughts about or felt like harming myself' and 0.28 with the item 'I have made plans to end my life' (both significant at 0.01 level).

However, for the SMI reference group the correlation between the practitioner judgement of Risk to others and the single item 'I have had thoughts about or felt like harming someone' decreased to 0.088 (non-significant). As a check, the correlation with the two-item self-reported risk to others score (HRUOTH) was also calculated at 0.049 (non-significant at 0.05 level).

When the above analyses were re-run for the CMH reference group, the correlation between the practitioner judgement of Suicide risk and the self-reported risk to self score (HRUSELF) decreased from 0.29 to 0.20 (significant at 0.05 level). Correlations with individual How Are You? items decreased to 0.19 for both items (both non-significant).

For the CMH reference group the correlation between the practitioner judgement of Risk to others and the single item 'I have had thoughts about or felt like harming someone' increased from 0.16 to 0.27 (significant at 0.01 level).

Analysis by reference group therefore had an interesting effect: for the SMI group it increased the correlation on the Risk to self items but reduced it on the Risk to others items; and vice versa for the CMH group. The increased correlation of nearly 0.4 on the Risk to self items in the SMI group might be considered especially significant, particularly in the context of the overall low patterns of correlation and the general emphasis on Risk to others within this group.

9.5 Agreement on the presence of risk within the 'at risk' group

In order to bring the above findings further into focus, it was decided to concentrate on the 'at risk' group.

There were 32 service users who were judged by practitioners to present a significant risk to others or a risk of suicide – 21 men and 11 women. One male service user was judged to be at risk on both counts. Of these, 24 belonged to the SMI reference group and eight to the CMH reference group. Within this group the correlation between the self-report items relating to self-harm and the judgement of suicide risk was 0.51 (significant, p=0.004), up from 0.29 for the population as a whole. However, there was no correlation between the item 'I have had thoughts about or felt like harming someone' and judgement of risk to others, Pearson correlation=0.097 (p=0.62).

The above analysis defined the 'at risk' group as being those service users judged by practitioners to present a significant or greater risk. An alternative approach is to define a service user as 'at risk' if *either* the practitioner or the service user consider a risk to be present. Since service users were not asked this directly, a threshold for the presence or absence of risk on How Are You? was created, thus permitting examination of levels of categorical 'agreement' between service user and practitioner.

In total, 28 service users scored >3 on the HRU risk to self items and 13 service users were judged by practitioners to be at significant risk or greater of suicide. Of the latter, five scored greater than 3 on the HRU items. Thus, less than one in five of the service

users with a high self-reported risk to self were identified by practitioners as posing a significant risk or greater of suicide. If 'agreement on the presence of suicide risk' between practitioner and service user is defined as the presence of both a practitioner rating of significant risk or greater and service user score of greater than 3 on the relevant How Are You? items, then of the 36 service users who were above threshold on either by the practitioner or self-report measure there was agreement on only five – a percentage agreement of less than 15%.

In total, 16 service users scored >2 on the HRU risk to others items whilst 18 service users were judged to be at significant risk or greater to others. Of the latter, 14 scored 0 on the HRU risk to others item and a further two scored 1. Thus, only two of the 18 reported thoughts of harming others more than occasionally (both of these scored 4, indicating 'most or all of the time'). Of the 16 service users who reported that they had thoughts of harming others often or most or all of the time, only two were identified as a risk to others by practitioners. If 'agreement on the presence of risk to others' between practitioner and service user is defined as a practitioner rating of significant risk or greater and a service user score of greater than 1 on the How Are You? item, then of the 32 service users who were above threshold on either measure there was joint agreement that risk was present in only two cases – a percentage agreement of about 6%. Even if the threshold for self-report is set at the highest level possible, including only service users who report thinking about harming others 'most or all of the time' there was agreement on only two out of nine service users – a percentage agreement of about 22%.

There thus appears to be a very low concordance between practitioner and service user perceptions of risk within the group of service users most at risk. Given that in routine practice, practitioner perceptions of risk are at least in part based upon service user self-report, this is a surprising finding and perhaps casts some light on the practical difficulty of identifying level of risk within the 'at risk' group.

9.6   Relationship between practitioner and service user perceptions amongst service users judged to be at risk of suicide

Correlations between practitioner judgements and self-report scores were explored amongst the group of service users judged to be at risk of suicide. Such analysis was somewhat tentative, given the very small number (13). The correlation between the item 'I have had thoughts about or felt like harming myself' and the practitioner judgement of risk of suicide was 0.35 (non-significant, p=0.24). However, the correlation between this item and the self-report item 'I have made plans to end my life' was 0.70 (significant, p=0.008), even higher than that found in the population as a whole. Thus, even in a group consistently self-reporting intense suicidal ideation the correlation with practitioner judgements of suicidality was low.

Paradoxically, however, within this sub-group the correlation between practitioner judgement of risk to *others* and self-report of thoughts of harming others was high, approaching significance (0.56, p=0.056).

9.7 Relationship between practitioner and service user perceptions amongst service users judged to be a risk to others

Correlations between practitioner judgements and self-report scores were also explored amongst the group of service users judged to be at risk to others. The correlation between the item 'I have had thoughts about or felt like harming someone' and the practitioner judgement of risk to others was 0.29 (non-significant, p=0.24). However, the correlation between this item and the practitioner judgement of suicide risk was 0.52 (significant, p=0.029). This suggests that self-report of thoughts of harming others may be a useful indicator of suicidality.[9]

Additionally, within this group the correlation between the self-report item 'I have had thoughts about or felt like harming myself' and the practitioner judgement of suicide risk was 0.55 (significant, p=0.016). The level of statistical significance in relation to these two items is all the more striking given the small numbers of service users involved.

However, it was striking that within this sub-group the correlation between practitioner judgement of risk to self and the item 'I have made plans to end my life' was negative at -0.12 (non-significant).

It was also striking that the correlation between the two self-report items 'I have had thoughts about or felt like harming myself' and 'I have made plans to end my life' was very low within this sub-group (0.14, non-significant). Thus, although in the population as a whole the correlation between these two items was high (0.56), within the sub-group judged to be most at risk to others there was almost no correlation.

A final set of analyses examined correlations between self-report items and practitioner judgements in members of the SMI reference group judged to present a risk to others. This produced a number of striking findings. Within the SMI group (n=12) the correlation between practitioner judgement of risk to others and the self-report item 'I have had thoughts about or felt like harming someone' was negative at −0.23 (p=0.50). However, the correlation between the item 'I have had thoughts about or felt like harming someone' and the item 'I have heard voices that frighten or upset me' was 0.70 (significant, p=0.01). Thus despite high levels of consistency in self-report there was no correlation with practitioner judgement.

Within the SMI group the correlation between the item 'I have had thoughts about or felt like harming someone' and practitioner judgement of suicide risk was a remarkable 0.69 (significant, p=0.013).

Also within the SMI group the correlation between practitioner judgement of risk to self and self-report of thoughts of self-harm was 0.68 (significant, p=0.01) but the correlation between practitioner judgement of suicide risk and the item 'I have made plans to end my life' was negative at -0.40 (non-significant).

---

[9] In the population as a whole the self-report risk to others score also correlated more highly with the practitioner judgement of Suicide risk (0.20, significant at 0.01 level) than it did with the practitioner judgement of Risk to others.

These findings coupled with those of the previous section suggest that there is a complex set of relationships between judgements of risk to self and of risk to others, both within and across practitioners and service users. This issue was therefore explored further.

9.8 Relationship between judgements of risk to self and risk to others

In 7.6 (final section) it was observed that the correlation between practitioner judgements of risk to self and risk to others differed according to which sub-group of service users was being looked at. Whilst in general terms an increase in perceived suicide risk is moderately correlated with an increase in perceived risk of harm to others, within the group who were perceived to pose a significant risk to themselves or others, the two types of risk were strongly *negatively* correlated.

The present data enable us to see whether there is a similar change in correlation in service users' perceptions. Specifically, the notion that risk of suicide decreases with increased risk of harm to others (and vice versa) appears to be disconfirmed by consideration of a similar table for the self-report items (Table 23, last column). Here it can be seen that the correlation between self-report of thoughts of self-harm and of thoughts of harming others is **greatest** in those judged to pose a risk.

Furthermore, whereas within the SMI group there was no correlation between practitioner judgement of risk to self and risk to others, there was a correlation of 0.46 on the self-report data. Correlations for the CMH group were similar.

Table 23: Correlations between practitioner and service user perceptions of risk to self and risk to others		
Population	**Practitioner judgement of suicide risk and risk to others**	**Service user report of thoughts of harming others and of harming self**
All (n=331)	0.22 (p<0.0001)	0.43*
All SMI (n=211)	0.054 (p=0.44)	0.38*
All CMH (n=104)	0.40 (p<0.001)	0.34*
All judged to be a risk (n=32)	-0.79 (p<0.0001)	0.51*
All judged not to be a risk (n=298)	0.15 (p=0.009)	0.46*
All judged to be a risk to others (n=20)	-0.77	0.52*
All judged to be a risk to others (n=13)	-0.41	0.22

More generally, it can be seen that service users are fairly consistent in their reporting of risk to self and others in the sense that there is a consistent positive correlation between the two sets of ratings. This contrasts dramatically with the sub-group dependence of the same judgements on the part of practitioners.

There would therefore seem to be a conflict between a very strong judgemental tendency on the part of practitioners and what is actually reported by service users, especially in relation to the SMI population. This leaves an acute dilemma: either practitioners are labouring under a misconception when making judgements of risk or

the type of self-report data reported upon in this study should not be interpreted as an indicator of risk to self or others.

This dilemma is at its most acute in relation to service users judged to pose a risk to others. A possible explanation is therefore that practitioners faced with a service user apparently presenting a risk to others focus on that element at the expense of consideration of possible risk to self. More generally, it may be that practitioners focus on assessment of the most acute risk at the expense of consideration of collateral risk. Thus it is possible that a strong presentation of suicidality results in lack of adequate consideration of possible risk to others.

9.9 Relationship of subsequent incidents to self-report scores

Two hundred and forty-one service users completed the HRU questionnaire. Of these, 24 were involved in a subsequent incident. Of the latter, 12 (50%) scored >1 and nine (37%) scored 5 or more on the HRU risk score. The mean HRU risk score was 3.3. In the population who were not involved in a subsequent incident, 38% scored >1 on the HRU risk score and about 12% scored 5 or greater. Thus a score of 5 or greater on the self-report measure resulted in a trebling of likelihood of subsequent involvement in a serious incident. This is a comparable level of enhanced predictiveness to that observed in high practitioner judgements of risk (see S.8.4).

There were 23 service users who had BOTH a practitioner rating of greater than 1 on risk to self or others AND a HRU score of greater than 1 on one of the HRU risk variables. Of these, five were involved in a subsequent incident. This rate of occurrence of subsequent serious incidents of 22% was comparable to the rate of those service users scoring greater than 1 on the practitioner rating. Thus, inclusion of this relatively low self-report score in conjunction with the practitioner judgement did not increase predictiveness.

9.10 Discussion

The main findings may be summarised as follows:

1) Practitioners are consistent in their approach to judgements of risk (see Chapter 7).

2) Service users are consistent in their self-reporting of ideas of harm to self and others.

3) There is limited agreement between practitioner judgements and service user perceptions of risk to self or others.

4) There is a complex relationship between practitioner judgements of risk to self and risk to others; and between such judgements and the corresponding self-report items.

5) Practitioners' and service users' judgements of risk are equally predictive of subsequent involvement in serious incidents (even though they do not agree with each other).

Before reviewing these findings, it is important to consider methodological and practical issues which may have affected the data. First, it could be argued that the interpretation of the self-report ratings leaves something to be desired. How true is it that a service

user who reports thinking about harming themselves or others is correctly interpreted as reporting themselves to be a risk to self or to others? Or, how true is it that someone who reports no such thoughts is indicating that they do not consider themselves to be a risk? The data offers no answers to these questions. However, whilst it may be the case that alternative phrasing of self-report questions could better capture concepts of risk, it is hard to see why the interpretation proffered here should be far from the truth: if someone reports that they are consistently planning to end their life, it defies logic not to interpret this as an indicator of suicidality (and this interpretation is backed up by other research, e.g. using the Beck Depression Inventory). Secondly, it is plausible that some individuals with definite plans to harm themselves or others would not admit to such thoughts on a self-report questionnaire, whether through a wish to conceal the fact, lack of insight or other reasons. This would partly explain the low level of agreement between self-report and practitioners, especially in cases where the practitioner suspects the service user of posing a risk, based on past history and clinical experience, whilst the service user denies it. Thirdly, the nature of the judgements being made by practitioner and service user may be fundamentally different. As evidenced in the previous chapter, practitioners appear to take into account the past when assessing risk of future behaviour. It is less clear whether this is likely to be the case with a service user, especially when How Are You? specifically asks them to report on their current state as opposed to making a judgement that includes consideration of their past behaviour.

These are all important methodological, practical and conceptual issues that would warrant further exploration. However, they would if anything be expected to mask or distort relationships rather than produce highly specific and striking findings.

Turning then to the findings, the key finding is the lack of agreement between service user and practitioner perspectives, despite the high levels of internal consistency (in general terms) demonstrated by both. Comparing the results on the risk to others and suicide risk items, it is notable that the level of practitioner-service user agreement is higher in the case of suicide risk than that of risk to others. However, at 15% compared to 6% within the 'at risk' population, it remains very low. The explanation for this is unclear. One possibility is that only a certain proportion of service users are willing to admit to thoughts of harming themselves or others, but that if they do they do so consistently. Alternatively, it might be that practitioners only assign a high weighting to self-report if other factors are also present, for example a history of risk behaviour or observed current behaviour. Whatever the explanation, a central issue is what weight should be given to the service user and practitioner perspectives. The data can be viewed from both sides. Trusting to the accuracy of practitioner judgement – which as described in previous chapters does seem to correlate reasonably with the presence of factors known to affect risk – the findings could be said to cast some doubt on the value of using self-report measures, such as How Are You? for the detection of risk, since many service users scored low on the self-report 'risk' items whilst apparently presenting a risk when assessed by practitioners. The opposite could also be argued: the finding that significant numbers of service users score highly on the How Are You? risk items in the absence of practitioner judgement of risk suggests a need to use self-report tools to enhance practitioner detection.

Additionally, analysis of the self-report measure shows that a substantial proportion of service users considered by practitioners to be a risk will not report themselves as having thoughts of suicide or of harming others. It is clear therefore that it would be

unwise to rely solely on such measures for detection since this would result in a significant proportion of false negatives. The unknown in both cases is whether the service users so detected (or not) in fact present a risk. This is hard to ascertain in the absence of a longitudinal study that tracks what does in fact subsequently occur – and even this would be hard to justify on ethical grounds since it would require withholding the How Are You? information indicating risk from the practitioners in case their subsequent actions affected later outcome.[10]

The How Are You? data confirmed the internal consistency of service user self-report. This applied to both the items on risk to self and risk to others considered separately, and the relationship between self-report of thoughts of harming others and of harming themselves. The latter was consistent at around 0.4 or higher in all sub-groups. It is therefore all the more disconcerting to find the correlation between practitioner judgements of risk to self and risk to others varying remarkably according to sub-group. As previously discussed (chapter 5) what appears to be occurring is that having identified a major risk, practitioners focus on that to the exclusion of other possible risks. If this is the case, then the self-report data would lend credence to the view that in so doing they are systematically underestimating these other risks.

This and other findings also further underscore the importance of reference groups in interpretation of risk data. Difference in reference group response was not along the lines that might have been expected. The SMI group was more consistent in self-report of risk to self than the CMH group and there was a higher correlation between reports of suicidal ideation and practitioner judgement of risk to self in the SMI reference group, despite the fact that in routine practice the latter are less likely to be given self-report questionnaires that include such items, such as the Beck Depression Inventory. Equally remarkable, within the SMI group there was a high level of consistency in self-report of risk to others but this correlated *negatively* with practitioner perceptions of risk to others.

Overall, then, the empirical findings suggest that we do not as yet fully understand either the relationship between practitioner assessment of risk and service user assessment of risk or the nature of practitioner judgements of risk. The best that can be said is that practitioner assessment appears to be reasonably in line with evidence-based practice, in that there is a reasonable correlation between practitioner ratings of level of risk and the presence of known risk factors. However, the extent to which this results in an accurate assessment of actual risk remains not only an unknown but is highly questionable. Similarly, service users are consistent in self-report but the extent to which such reports should be used by practitioners for risk management planning purposes is unclear.

These conclusions have several implications. First, there is an urgent need for a longitudinal study of risk which includes repeat measures from both practitioner and

---

[10] An indication might be gleaned through examination of the characteristics of those service users scoring highly on How Are You? but not rated as a suicide risk by the practitioners. It might be predicted, for example, that there would be a higher proportion of new service users in this group – the absence of background information and previous occurrences of suicidal behaviour on the person might then provide an explanation as to why the practitioner ratings are low. In fact, however, this was disconfirmed: the proportion of 'new' service users, defined as those new to the service in the past year was about 12%, similar to the figure for the population as a whole.

service user perspective, and also includes accurate recording of subsequent incidents. Given the relatively modest numbers of individuals assessed as being a significant risk to themselves or others (31 of the 492 service users who received the Risk Profile and completed How Are You?), such a study is likely to have to draw upon a very large sample.

A second implication is that at this stage it is probably unwise to rely solely upon either practitioner-based assessment tools or self-report measures as methods of detecting risk. Both would appear to be useful supports to clinical practice. It would appear that self-report measures such as How Are You? detect many more individuals who have frequent thoughts of harming themselves or others than would ordinarily be detected by clinicians. Whether these service users actually pose a risk may be unknown, but it would seem to be unwise to discount such information if it is available. The results thus further illustrate the importance of including both practitioner and service user perspectives in any approach to assessment.

Finally, the results demonstrate the capacity of relatively simple routinely-collected data to throw up findings of scientific and clinical significance – findings which might never have come to light in a research context. As such they provide a compelling demonstration of the power of the infometric approach and the potential for using findings derived routinely from infometric tools to advance our understanding.

10 Elements of Effective Risk Management

This final section reviews the implications of the research and development described for the day-to-day management of risk. Whilst findings based upon such a modest sample can only be provisional, the data are highly suggestive in a number of areas.

10.1 The Risk Profile

The Risk Profile would appear to be a good candidate for routine use as a standardised method of recording risk factors, risk judgements and risk outcomes (such as changes in level of risk, subsequent involvement in serious incidents). The tool appears to have strong psychometric properties, leading to consistent judgements of risk. The scoring systems are internally consistent and compare favourably to other established instruments; and the tool is sensitive to change. If completed correctly, the tool provides some guarantee that relevant risk factors are adequately considered during the process of risk assessment; and of course the completed tool provides evidence of what risk factors were or were not identified; and of what judgements of risk were made based upon these and broader clinical considerations. Furthermore, the standard format enables easy communication of such judgements and the basis for them to other members of the team than those involved in the direct assessment.

Additionally, the structure of the tool lends itself to exploratory analysis of aggregated data. The clear separation of judgements regarding the presence or absence of risk factors from the actual judgements of perceived risk enables the relationship between the two to be investigated and modelled. Similarly, the recording of two types of 'risk outcome' – changes in perceived levels of risk and subsequent involvement in a serious incident – enables a modelling of the relationships between risk factors, risk judgements and subsequent outcomes, as well as the tracking of change risk status of individuals.

10.2 Investigating risk

The analysis produced a number of suggestive findings concerning the nature of judgements of risk and their accuracy, some of which have significant practical consequences. The first of these is that judgements of perceived risk would seem in fact to comprise two separate judgements:

(1) A judgement as to whether 'risk' is present to a significant degree.

(2) If the risk is judged to be present, a further judgement as to the seriousness and immediacy of that risk.

This distinction implies, for example, that it would be perfectly possible to develop an accurate model of indicating whether a risk is present or absent, without having any method of predicting the seriousness of that risk; or to put it more positively, it suggests that two models are required to 'predict' or 'assign' risk: one model to determine whether a service user presents a risk in a particular category; and another to indicate the seriousness of the risk. This is a departure from the approach to modelling in the risk literature, perhaps because it has primarily been concerned with groups of service users, such as forensic service users, who by definition are already known to present a risk.

Finally, within the 'high risk' group there was a surprisingly high *negative* correlation between practitioner judgements of risk to self and others in those with a high risk, even

though within the population as a whole the two types of risk were positively correlated. This underlines the need to distinguish between the two types of risk judgement and suggests that clinicians naturally focus on the most obviously apparent risk, perhaps at the expense of other risks. If correct, this has important training implications.

10.3 Prediction

This leads us to the complex issue of 'prediction'. As previously observed, most clinicians will say it is impossible to predict risk behaviour – at the same time as asking whether the tool does so. Before considering whether it can or it can't a more general question needs to be asked, which is whether 'prediction' is really the appropriate goal of a risk assessment tool – in any stronger sense than that a service user rated as 'high risk' is more likely to be involved in a serious incident than someone rated as low or no risk. A major finding was that although those assessed as high risk initially were more likely to be involved in a subsequent incident, the majority of service users who upon follow-up assessment had been involved in a serious incident had **not** been regarded as high risk in their initial assessment. It is possible that the perceived risk status of such service users did change before their follow-up assessment but that this simply had not been recorded on the form. In this sense, the nature of the study, which was such that risk status could not be updated as required, might well have led to an underestimate of the true predictiveness of perceived high risk status. Without further research it is difficult to know if this is true or not. Similarly, it is probable (one would hope) that preventive action was taken with service users perceived as high risk. Assuming that such action has some effectiveness, this would again result in an apparent underestimation of the predictiveness of initial judgements of risk. Whilst these are valid points, they don't invalidate the basic finding that initial perceived risk status was at best moderately predictive; and in the absence of more recording of risk status, which may well be difficult to achieve, we are left with a situation where prediction at the individual level may well not be a realistic goal.

The potential benefits of a predictive model are further undermined by the reflection that, from a practical utility standpoint, a predictive model is only useful if it offers a level of certainty that **_adds_** to clinical decision-making. In many cases it is obvious that someone presents a risk (e.g. if someone has a repeated history of suicide attempts and is indicating they are likely to try again) and the presence of a predictive model that confirms this would not add much to clinical practice. If these cases are removed from the equation then predictive models become far less accurate.

Such considerations suggest that 'prediction' in the sense of predicting the likelihood of any specific _individual_ becoming engaged in risk behaviour, may not be a sensible goal. However, it might well be that group prediction is a more practically beneficial goal. For example, if the average levels of risk across caseloads differ across two otherwise similar clinical teams it might well be sensible to deploy more staff to the team with the higher average risk level. The data on risk outcomes and risk replacement rates provides a glimpse (however provisional) of the naturally-occurring 'life cycle' of risk in mental health services – a cycle in which some risks reduce whilst new ones emerge. Based upon this life cycle and an estimate of the resources required to manage different levels of risk, it might be possible to more rationally allocate resources to manage both actual and emerging risks. This would depend both upon the 'life cycle' having a reasonable degree of consistency, and upon a better understanding of the life cycles of risk in

different settings or populations. Nevertheless, there are the beginnings here of a promising and powerful methodology, based upon the degree of predictability not of individual behaviour (which seems a poor candidate for prediction) but rather of population behaviour (which seems a better candidate).

10.4 Individual vs population-based risk strategies

A broader question is the appropriate balance between focusing risk management resources upon individual 'high-risk' service users and population-based risk management strategies. One in five individuals was perceived to present a significant risk to self or others. It would seem unlikely that resources would ever permit special measures to be taken with 20% of the clinical population.

Thus even on the assumption that initial identification of risk is accurate, practitioners are then faced with the further task of identifying which of those 20% are of greatest risk and can be apportioned greater resources to aid risk prevention. The complexity of such decisions in the absence of a standard tool is emphasised by the number of risk factors: those judged to present a significant or greater risk averaged about 17 risk factors. This must make it difficult for practitioners to distinguish (say) between someone with 13 risk factors who is not a significant risk and someone with 18 factors who is. A standard tool does not resolve the difficulty but it does least provide a balanced view and make both quantitative and qualitative differences in risk factors easier to observe.

The appropriateness of a population-based approach is also indicated by the need to distinguish between 'reference groups' when considering aggregated data: a model of risk management that may be reasonable for people with severe mental illness may be quite inappropriate to those with other types of mental health problems.

Finally, the need for a group-based perspective on risk is also highlighted by the poor correlations between clinician and service user perceptions of risk. Overall, there was a low correlation between service user and practitioner judgements of risk, but the two sets of data were equally predictive. This suggest that there are two distinct populations (at least), one for whom the Risk Profile is moderately predictive and another for whom service user self-report is moderately predictive. Whether it is possible to identify a priori who belongs to which group is a question that was not addressed, but one that would clearly be useful to explore further.

10.5 Conclusions

In developing the Risk Profile, the aim was not just to develop a 'risk assessment tool', but to develop a tool that will support day-to-day risk management. Whilst an assessment tool is an important first step in risk management, more important are the broader strategies that it can support and inform. This initial research suggests that as well as its acceptability in routine practice, Risk Profile data can generate practical insights that can feed into training, clinical decision-making and resource management. Going forward, there is the opportunity to collate much larger-scale data from electronic health records. This could provide the basis for external as well as internal comparisons and benchmarking, as well as informing more sensitive analysis in this complex area.

11 References

Allen, J. (1997). Assessing and managing risk of violence in the mentally disordered. J. Psychiatric Mental Health Nursing 1997, Oct 4(5): 369:378.

Appleby, L., Shaw, J., Amos, T., McDonnell, R., Harris, C., McCann, K., Kiernan, K., Davies, S., Bickley, H., Parsons, R. (1998). Suicide within 12 months of contact with mental health services: national clinical survey. BMJ Volume 318, 8th May 1999 pp.1235-1239.

Beck, A.T. and Kouvacs, M. (1979). Assessment of Suicidal Intention: The Scale for Suicidal Ideation. Jnl of Consult. and Clin.Psychol. 1979, Vol 46(2), 343-352.

Bindman, J., Beck, A., Thornicroft, G., Knapp, M., and Szmukler, G. (2000). Psychiatric service users at greatest risk and in greatest need. Impact of the Supervision Register Policy. The British Journal of Psychiatry (2000) 177: 33-37.

Clifford, P. (1997) The Core Assessment & Outcomes Package for Mental Health Services V.1. CORE, UCL, London.

Clifford, P. (2001). The National Outcome Audit of Severe Mental Illness. Report to the Department of Health. CORE, UCL, London.

Clifford, P., Katsavdakis, K., Lyle, J., Fultz, J., Allen, J., Graham, P. (2002). How Are You? Further development of a generic quality of life outcome measure. Journal of Mental Health.

Clifford, P. (2003) An Infometric Approach to Mental Health Outcome Measurement. Ph.D. thesis, UCL, London.

Clifford, P. (2005) The Core Assessment & Outcomes Package for Mental Health Services V.5. FACE Recording & Measurement Systems. Nottingham.

Department of Health (2000). Effective Care Co-ordination in Mental Health Services: Modernising the Care Programme Approach. Department of Health. HMSO.

Dolan, M., Doyle, M., (2000). Violence risk prediction. Clinical and actuarial measures and the role of the Psychopathy Checklist. The British Journal of Psychiatry (2000) 177: 303-311.

Douglas, K.S., Ogloff, J.R., Nicholls, T.L. (1999). Assessing risk for violence among psychiatric service users: the HCR-20 violence risk assessment scheme and the Psychopathy Checklist: Screening Version. J Consult Clin Psychol 1999 Dec;67(6):917-30.

Eastman, N. (1997). The Mental Health (Service users in the Community) Act, 1995. British Journal of Psychiatry, 170, 492-496.

Edwards, A., Matthews, E., Pill, R. (1998). M Fam Pract 1998 Aug;15(4):301-7.

Ferris, L.E., Sandercock, J., Hoffman, B., Silverman, M., Barkun, H., Carlisle, J., Katz, C. (1997). Risk assessments for acute violence to third parties: a review of the literature.

Can J Psychiatry 1997 Dec;42(10):1051-60.

Fuller, J. and Cowan, J. (1999) Risk assessment in a multidisciplinary forensic setting: Clinical judgement revisited. Journal of Forensic Psychiatry, 10, 276-289.

Hall, R.C. and Platt, D.E. (1999). Suicide risk assessment: a review of risk factors for suicide in 100 service users who made severe suicide attempts. Evaluation of suicide risk in a time of managed care. Psychosomatics 1999 Jan-Feb;40(1):18-27.

Hammond, S.M. and O'Rourke, M.M. (1997). Developing a Psychometric Model of Risk Assessment: the case of the RAMAS.

Hare, R.D. (1980). A Research Scale for the assessment of psychopathy in criminal populations. Personality and Individual Differences, 1, 111-117.

Lingham, R., Candy, J. and Bray, J. (1996). Report of the Inquiry into the Treatment and Care of Raymond Sinclair. Aylesford: West Kent Health Authority.

Linke, S. (1997). Assessing and Managing Suicide Risk. Coster, S. and Clifford, P. (eds). CORE Miniguide Series, No. 1. British Psychological Centre for Outcomes, Research and Effectiveness, University College London.

Linke, S. (1998). Assessment and Management of Service users Presenting Risk to Others. Coster, S. and Clifford, P. (eds). CORE Miniguide Series, No. 2. British Psychological Centre for Outcomes, Research and Effectiveness, University College London.

McNiel, D.E. and Binder, R.L. (1995). Correlates of accuracy in the assessment of psychiatric inservice users' risk of violence. Am J Psychiatry 1995 Jun;152(6):901-6.

McNiel, D.E., Sandberg, D.A. and Binder, R.L. (1998). The relationship between confidence and accuracy in clinical assessment of psychiatric service users' potential for violence.

Law Hum Behav 1998 Dec;22(6):655-69.

McNiel, D.E., Lam, J.N. and Binder, R.L. (2000) Relevance of interrater agreement to violence risk assessment. J Consult Clin Psychol 2000 Dec;68(6):1111-5.

Menzies, R., Webster, C., McMain, S. et al. (1994). The dimensions of dangerousness revisited: Assessing forensic predictions about violence. Law and Human Behaviour, 18, 1-28.

Monahan, J. and Steadman, H. (1996). Toward a Rejuvenation of Risk Assessment Research. In Violence and Mental Disorder, editors: Monahan, J. and Steadman, H., Chicago: University of Chicago Press.

Monahan, J., Steadman, H.J., Appelbaum, P.S., Robbins, P.C., Mulvey, E.P., Silver, E., Roth, L.H. and Grisso, T. (2000). Developing a clinically useful actuarial tool for assessing violence risk. Br J Psychiatry 2000 Apr;176:312-9.

Morgan, H.G. and Owen, J.H. (1990). Persons at Risk of Suicide: Guidelines on Good Clinical Practice. The Boots Company PLC.

Morgan, H.G. and Stanton, R. (1997). Suicide among psychiatric in-service users in a changing clinical scene. Suicidal ideation as a paramount index of short-term risk. Br J Psychiatry 1997 Dec;171:561-3.

Munro, E. and Rumgay, J. (1997). The role of risk assessment in reducing homicides by people with mental illness. (Manuscript submitted to British Journal of Psychiatry).

NHS Executive (1994). Health Service Guidelines HSG (94) 27. London: Department of Health.

O'Rourke, M.M. (1995). The Risk Assessment and Management Schedule: Users Manual. Surrey: Heathlands Mental Health Trust.

O'Rourke, M.M., Hammond, S.M. and Davies, J. (1997). Risk Assessment and Risk Management: the way forward. Psychiatric care, 4, 132-138.

O'Rourke, M.M. and Hammond, S. (2000). Risk Management: Towards Safe, Sound, and Supportive Service. March 2000. Surrey Hampshire Borders NHS Trust.

Shaw, J., Appleby, L., Amos, T., McDonnell, R., Harris, C., McCann, K., Kiernan, K., Davies, S., Bickley, H. and Parsons, R. (1999). Mental disorder and clinical care in people convicted of homicide: national clinical survey. BMJ Volume 318, 8th May 1999 pp.1240-1244.

Steadman, H., Monahan, J., Robbins, P., Appelbaum, P., Grisso, T., Klassen, D., Mulvey, E.P. and Roth, L. (1993). From dangerousness to risk assessment: Implications for appropriate research strategies. In Hodgins, S. (Ed.) Mental Disorder and Crime. London: Sage.

Steadman, H., Monahan, J., Appelbaum, P. et al, (1996). Designing a New Generation of Risk Assessment Research. In Violence and Mental Disorder, editors: Monahan, J. and Steadman, H., Chicago: University of Chicago Press.

Stedman, T., Yellowlees, P., Mellsop, G., Clarke, R. and Drake, S. (1997). Measuring Consumer Outcomes in Mental Health. University of Queensland, St. Lucia, QLD.

Stein, L.I. and Test, M.A. (1978). An Alternative to Mental Hospital Treatment. New York. Plenum Press.

Stewart, A.L. and Ware, J.E., Editors (1992). Measuring Function and Well-Being. The Medical Outcomes Study Approach. Duke University Press, Durham and London, 1992.

Swanson, J., Swartz, M., Estroff, S., Borum, R., Wagner, R. and Hiday, S. (1998). Psychiatric impairment, social contact, and violent behavior: evidence from a study of outpatient-committed persons with severe mental disorder. Soc Psychiatry Psychiatr Epidemiol 1998 Dec;33 Suppl 1:S86-94.

Taylor, P. and Gunn, J. (1999) Homicides by people with mental illness: myth or reality. British Journal of Psychiatry, 174, 9-14.

Wing, J., Curtis, R. and Beevor, A. (1996). Health of the Nation Outcome Scales: Report on research. London: Royal College of Psychiatrists.

Wright, J.G. and Feinstein, A.R. (1992). A comparative contrast of clinimetric and psychometric methods for constructing indexes and rating scales. J Clin Epidemiol, Vol. 45, No. 11, pp.1201-1218, 1992.