Implementation of treatment guidelines for specialist mental health care of severely mentally ill patients

Markus Koesters[1], Francesca Girlanda[1,2], Esra Ay[1], Andrea Cipriani[2], Corrado Barbui[2]

[1]Division Psychiatry ll, Ulm University, Guenzburg, Germany

[2]Department of Public Health and Community Medicine, Section of Psychiatry and Section of Clinical Psychology, University of Verona, Verona, Italy

Contact person

Dr. Markus Kösters
Division Psychiatry II, Ulm University
Bezirkskrankenhaus Guenzburg
Ludwig-Heilmeyer-Str. 2
D-89312 Guenzburg/Germany
Phone: ++49-8221-96-2869

Background

Description of the condition

In the last few decades, guideline development and evaluation methods have been significantly improved, resulting in a considerable number of evidence-based mental health guidelines in many countries across the world. Evidence-based practice guidelines in psychiatry are viewed as an essential asset if appropriately developed and applied (Lehman 2006). Nevertheless, these evidence-based interventions are not easily translated into practice, leading to a gap between what is known and what is done Berwick 2003). Evidence-based recommendation makes little sense if it is not translated into action (Barbui 2011) and one approach to improve the translation of evidence into practice is the dissemination and implementation of clinical guidelines.

While the pathway from evidence generation to evidence synthesis and guideline development is highly developed and quite sophisticated, the pathway from evidence-based guidelines to an evidence-based practice is much less developed. Available evidence is scant and inconclusive.

Only few reviews have covered the topic of guideline implementation in severe mental disorders (Weinmann et al., 2007; Morriss, 2008; Drake et al., 2009). Implementation methods range from

simple interventions, such as dissemination of educational material, to more complex and multifaceted interventions, including tutorial and consultation sessions, use of treatment algorithms, reminder systems, audit and feedback, and use of psychological theories to overcome obstacles (Grimshaw 2004). A review published in 2007 revealed that multifaceted interventions were more likely to have an impact on doctor performance and patient outcome, albeit effect sizes were generally modest (Weinmann 2007).

There is an urgent need to ascertain whether guidelines have an impact on doctor/practitioner performance and on patient outcomes, and to examine how guidelines should be implemented to maximise benefit at sustainable costs. This is particularly relevant for agencies involved in producing and delivering evidence-based recommendations, including international organizations such as the World Health Organization (WHO), scientific bodies such as the World Psychiatric Association or the American Psychiatric Association, national institutes such as the UK National Institute for Health and Clinical Excellence (NICE), but also for those with responsibilities in delivering high-quality mental healthcare, including national and local managers of mental health care systems, scientific organizations, or even single healthcare professionals.

Objectives

1. To summarize the evidence pertaining to the effects of guideline implementation on provider performance and patient outcomes in severely mentally ill patients.
2. To explore the performance of different strategies for guideline implementation.

Methods

Criteria for considering studies for this review

Types of studies

Although randomised controlled trials (RCTs) are considered the “gold standard” for the experimental evaluation of interventions, it has to be taken into account that guideline implementation is a complex task. Due to the difficulties in evaluating complex interventions these are frequently evaluated in non-randomized trials. As a consequence, reducing the evidence to RCTs may miss a substantial part of the evidence. In the present review, we therefore will include RCTs regardless of the randomisation level (patient, or health care provider), controlled clinical trials (CCT), controlled “before and after“ (CBA) studies and observational studies). To be comprehensive, studies

will be included regardless of study duration or quality of guideline implemented. We will include studies published in all languages.

Types of participants

Adults, however defined, with severe mental disorders, namely schizophrenia, psychosis, bipolar disorder, and severe depression will be included. The expected sample will be representative for patients treated in specialist mental health care settings.

We will not include studies in non-adult populations because of the differences in medical decision making in children and adolescents including the parents/guardian role. As we are interested in making sure that information is relevant to the care of severe mental disorders in specialist settings, only studies with participants recruited in mental healthcare settings will be included.

Types of interventions

We will include any active or passive guideline implementation strategy. We define guidelines braoadly as systematically developed statements (or algorithms, flow-charts, tables) to assist appropriate health care decisions for specific clinical circumstances. We define implementation as any planned process and systematic introduction of guidelines, the aim being that these are given a structural place in professional practice. Passive strategies, such as guideline distribution, will be included. Interventions will be classified according to a taxonomy developed by the Cochrane Effective Practice and Organisation of Care Review Group (EPOC) (http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/datacollectionchecklist.pdf).

The following comparisons will be included:

1. Guideline implementation strategy versus usual care ('no intervention' control)

2. Guideline implementation strategy A versus guideline implementation strategy B

Types of outcome measures

It is expected that outcomes will differ in different studies according to the characteristics and purposes of the guideline under scrutiny. Outcomes will be grouped into process outcomes (performance of healthcare providers) and patient outcomes for each condition. Studies that evaluate only satisfaction and/or knowledge of medical professionals will be excluded.

Primary outcomes

The following process outcome will be considered.

1. Practioner impact

As defined by each of the studies.

Secondary outcomes

The following patient outcomes will be considered.

1. Global state

1.1 Clinically significant response in global state - as defined by each of the studies.

2. Satisfaction with care

As defined by each of the studies.

3. Treatment adherence

As defined by each of the studies.

4. Drug attitude

As defined by each of the studies.

5. Quality of life

As defined by each of the studies.

'Summary of findings' table

We will use the GRADE approach to interpret findings and use GRADE profiler to create 'Summary of findings' tables (Guyatt 2011). These tables provide outcome-specific information concerning the overall quality of evidence from each included study in the comparison, the magnitude of effect of the interventions examined, and the sum of available data on all outcomes we rate as important to patient-care and decision making.

Search methods for identification of studies

Electronic searches

Medline, EMBASE, PsycLit, PSYNDEX, the Cochrane Database of systematic reviews (CDSR) and the Cochrane Central Register of Controlled Trials (CENTRAL) will be searched for relevant literature using a comprehensive search strategy. The Cochrane Schizophrenia Group Specialised Register will additionally be searched. It contains more than 11,000 coded studies (May 2009), and covers a number of databases that index scientific literature from almost every country of the world, with a specific attention to low- and middle-income countries and non-English literature (http://szg.cochrane.org/cochrane-schizophrenia-group-specialised-register).

An electronic search of the registers of the western databases will be run using the search strategy listed in the Appendix.

Searching other resources

Reference lists

As this review will also include non-randomized controlled trials the search will be supplemented by a careful hand search of the references of systematic reviews and studies included in the review.

Data collection and analysis

Selection of studies

Two review authors (FG, EA) will inspect all abstracts of studies identified as above and identify potentially relevant reports. In addition, to ensure reliability, CB and MK will inspect a random sample of these abstracts, comprising 30% of the total. Where disagreement occurs, this will be resolved by discussion, or where there is still doubt, the full article will be acquired for further inspection. The full articles of relevant reports will be acquired for reassessment and carefully inspected for a final decision on inclusion. Judgments will not be blinded to the names of the authors, institutions or journal of publication. Where difficulties or disputes arise, we will ask AC for help and if it is impossible to decide, we will add these studies to those awaiting assessment and contact the authors of the papers for clarification.

Data extraction and management

1. Data Extraction

Using a electronic form for data collection (EPIDATA), FG and EA will extract data from all included studies. To ensure reliability, MK and CB will independently extract data from a random sample of these studies, comprising 30% of the total. Again, any disagreement will be discussed, decisions documented and, if necessary, we will contact the authors of studies for clarification. With any remaining problems AC will help clarify issues and the final decisions will be documented. We will extract data presented only in graphs and figures whenever possible, but we will only include the data if two review authors independently have the same result. In case of missing data the authors of the primary studies will be contacted.

2. Data management

2.1 Scale-derived data

We will include continuous data from rating scales only if: a. the psychometric properties of the measuring instrument have been described in a peer-reviewed journal (Marshall 2000); and b. the measuring instrument is not written or modified by one of the trialists for that particular trial.

2.2 Endpoint versus change data

There are advantages of both endpoint and change data. Change data can remove a component of between-person variability from the analysis. On the other hand, calculation of change needs two assessments (baseline and endpoint) which can be difficult in unstable and difficult to measure conditions such as schizophrenia. We have decided primarily to use endpoint data and only use change data if the former are not available. Endpoint and change data will be combined in the analysis as we will use mean differences (MD) rather than standardised mean differences (SMD) throughout (Higgins 2008, chapter 9.4.5.2).

2.3 Skewed data

Continuous data on clinical and social outcomes are often not normally distributed. To avoid the pitfall of applying parametric tests to non-parametric data, we aim to apply the following standards

to all data before inclusion: when a scale starts from the finite number zero, the standard deviation(SD), when multiplied by two, is less than the mean (as otherwise the mean is unlikely to be an appropriate measure of the centre of the distribution (Altman 1996); if a scale starts from a positive value (such as the Positive and Negative Syndrome Scale (PANSS) which can have values from 30 to 210), we will modify the calculation described above to take the scale starting point into account. In these cases, skew is present if 2SD>(S-S min), where S is the mean score and S min is the minimum score. Endpoint scores on scales often have a finite start and end point and these rules can be applied. When continuous data are presented on a scale that includes a possibility of negative values (such as change data), it is difficult to tell whether data are skewed or not. We will enter skewed data from studies of less than 200 participants in additional tables rather than into an analysis. Skewed data pose less of a problem when looking at means if the sample size is large and we will enter such data into the syntheses.

2.4 Common measure

To facilitate comparison between trials, we will convert variables that can be reported in different metrics, such as days in hospital (mean days per year, per week or per month) to a common metric (e.g. mean days per month).

2.5 Conversion of continuous to binary

Where possible, efforts will be made to convert outcome measures to dichotomous data. This could be done by identifying cut-off points on rating scales and dividing participants accordingly into 'clinically improved' or 'not clinically improved'. It is generally assumed that if there has been a 50% reduction in a scale-derived score such as the Brief Psychiatric Rating Scale (BPRS, Overall 1962) or the PANSS (Kay 1986), this could be considered as a clinically significant response (Leucht 2005a; Leucht 2005b). If data based on these thresholds are not available, we will use the primary cut-off presented by the original authors.

2.6 Direction of graphs

Where possible, we will enter data in such a way that the area to the left of the line of no effect indicates a favourable outcome for implementation strategies.

Assessment of quality and risk of bias in included studies

Review authors FG and EA will independently assess the quality risk of bias of each included study with the Downs and Black instrument (REF: Downs & Black 1998 ??). The Black and Downs instrument contains 27 items covering the domains of reporting, external validity, bias, confounding and power. The instrument allows the assessment of the quality of randomised and non-randomised studies. Where inadequate details of trial characteristics are provided, we will contact the authors of the studies in order to obtain further information. Disagreements will be resolved by discussion with MK,CB and AC.

Measures of treatment effect

1. Binary data

For binary outcomes, we will calculate a standard estimation of the risk ratio (RR) and its 95% confidence interval (CI). It has been shown that RR is more intuitive (Boissel 1999) than odds ratios and that odds ratios tend to be interpreted as RR by clinicians (Deeks 2000). For statistically significant results, we will use 'Summary of findings' tables to calculate the number needed to treat to provide benefit /to induce harm statistic and its 95% CI.

2. Continuous data

We will analyse continuous data using mean differences (MD) (with 95% confidence intervals (CI)) or standardised mean differences (SMD) (where different measurement scales are used).

Unit of analysis issues

1. Cluster trials

Studies increasingly employ 'cluster randomisation' (such as randomisation by clinician or practice) but analysis and pooling of clustered data poses problems. They are commonly analysed as if the randomisation was performed on the individuals rather than the clusters. In this case, approximately correct analyses will be performed by dividing the binary data (the number of participants and the number experiencing the event) as presented in a report by a 'design effect' (Higgins 2011).This is calculated using the mean number of participants per cluster (m) and the intra-class correlation coefficient (ICC) [Design effect = 1+(m-1)*ICC] (Higgins 2011). If the ICC is not reported it will be assumed to be 0.1 (Ukoumunne 1999). For continuous data only the sample size will be reduced; means and standard deviations will remain unchanged.

2. Studies with multiple treatment groups

Where a study involves more than two treatment arms, we will present all relevant treatment arms in the comparisons. If data are binary, we will simply add and combine the data within the two-by-two table. If data are continuous, we will combine the data following the guidelines in the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2011). Where the additional treatment arms are not relevant, these data will not be reproduced.

Dealing with missing data

1. Overall loss of credibility

At some degree of loss of follow-up, data must lose credibility (Xia 2009). For any particular outcome should more than 50% of data be unaccounted for, we will not reproduce these data or use them within analyses. If, however, more than 50% of those in one arm of a study are lost, but the total loss is less than 50%, we will mark such data with (*) to indicate that such a result may well be prone to bias.

When binary or continuous outcomes are not reported, we will ask the study authors to supply the data.

2. Binary data

In the case where attrition for a binary outcome is between 0% and 50% and where these data are not clearly described, we will present data on a 'once-randomised-always-analyse' basis (an intention-to-treat analysis). Those leaving the study early will be considered to have the same rates of negative outcome as those who completed, with the exception of the outcome of death. We will undertake a sensitivity analysis to test how prone the primary outcomes are to change when 'completed' data only are compared with the intention-to-treat analysis using the above assumption.

When data on people who leave early are carried forward and included in the efficacy evaluation (Last Observation Carried Forward, LOCF), they will be analysed according to the primary studies; when these people are excluded from any assessment in the primary studies, they will be considered as having the negative outcome.

3. Continuous

3.1 Attrition

In the case where attrition for a continuous outcome is between 0 and 50%, and data only from people who complete the study to that point are reported, we will present and use these.

3.2 Standard deviations

For continuous outcomes, if standard deviations (SDs) are not reported, but an exact standard error (SE) and confidence intervals (CIs) are available for group means, and either the 'P' value or 't' value are available for differences in the mean, we will calculate them according to the rules described in the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2011). When only the SE is reported, we will calculate SDs by the formula SD = SE * square root (n) (Higgins 2011). The Cochrane Handbook for Systematic Reviews of Interventions present detailed formulae for estimating SDs from P values, t or F values, CIs, ranges or other statistics. If these formulae do not apply, we will calculate the SDs according to a validated imputation method which is based on the SDs of the other included studies (Furukawa 2006).

Assessment of heterogeneity

1. Clinical and methodological heterogeneity

Firstly, we will consider all the included studies to judge clinical and methodological heterogeneity, paying due attention to any differences in types of implementation strategies and outcome measures. If inspection of studies reveals considerable heterogeneity of guideline implementation strategies and outcome measures, formal meta-analyses will not be carried out. Any disagreement will be discussed and final decisions documented.

2. Statistical heterogeneity

3.1 Visual inspection

We will visually inspect graphs to investigate the possibility of statistical heterogeneity.

3.2 Employing the I^2 statistic

We will investigate heterogeneity between studies by considering the I^2 method alongside the Chi^2 'P' value. The I^2 provides an estimate of the percentage of inconsistency thought to be due to chance (Higgins 2003). The importance of the observed value of I^2 depends on i. magnitude and direction of effects and ii. strength of evidence for heterogeneity (e.g. 'P' value from Chi^2 test, or a confidence interval for I^2). We will interpret an I^2 estimate greater than or equal to 50% accompanied by a statistically significant Chi^2 statistic as evidence of substantial levels of heterogeneity (Section 9.5.2 - Higgins 2008). When substantial levels of heterogeneity are found in the primary outcome, we will explore reasons for heterogeneity (Subgroup analysis and investigation of heterogeneity).

Assessment of reporting biases

1. Protocol versus full study

Reporting biases arise when the dissemination of research findings is influenced by the nature and direction of results. These are again described in section 10.1 of the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2011). We will try to locate protocols of included trials. If the protocol is available, we will compare the outcomes in the protocol and in the published

report. If the protocol is not available, we will compare the outcomes listed in the methods section of the trial report with the reported results.

2. Funnel plot

We are aware that funnel plots may be useful in investigating reporting biases but are of limited power to detect small-study effects. We will not use funnel plots for outcomes where there are 10 or fewer studies, or where all studies are of similar sizes. In other cases, where funnel plots are possible, we will seek statistical advice in their interpretation.

Data synthesis

Data will only be combined within a diagnostic group.

As reported above (Assessment of heterogeneity), we will only calculate summary measures of intervention effect for studies assessing the impact of similar guideline implementation strategies and using similar outcome measures. If summary measures are calculated, we will employ a random-effects model for analyses throughout, as it takes into account differences between studies even if there is no statistically significant heterogeneity. The disadvantage of the random-effects model is that it puts added weight onto the smaller of the studies, that is those trials that are most vulnerable to bias. The reader is, however, able to choose to inspect the data using the fixed-effect model.

Subgroup analysis and investigation of heterogeneity

1. Subgroup analysis

Subgroup analyses are prospectively planned for:

Quality of implemented guideline

Guideline implementation strategy

Outcomes

2. Investigation of heterogeneity

If inconsistency is high, this will be reported. First, we will investigate whether data have been entered correctly.

Second, if data are correct, we will visually inspect the graph and remove outlying studies to see if homogeneity is restored. Should this occur with no more than 10% of the data being excluded, we will present the data. If not, we will not pool the data but discuss these issues.

Should unanticipated clinical or methodological heterogeneity be obvious, we will simply state hypotheses regarding these for future reviews or versions of this review. We pre-specify no characteristics of studies that may be associated with heterogeneity except the quality of the trial method. Should another characteristic of the studies be highlighted by the investigation of heterogeneity, perhaps some clinical heterogeneity not hitherto predicted but plausible causes of heterogeneity these post hoc reasons will be discussed and the data analysed and presented. However, should no reasons for the heterogeneity be clear, the final data will be presented without a meta-analysis. If data are clearly heterogeneous it may be misleading to quote an average value for the intervention effect.

Sensitivity analysis

No sensitivity analyses are planned.

Reference List

Berwick DM. (2003) Disseminating innovations in health care. JAMA 289:1969-75.

Bermejo, I., Schneider, F., Kriston, L., Gaebel, W., Hegerl, U., Berger, M. et al. (2009). Improving outpatient care of depression by implementing practice guidelines: A controlled clinical trial. International Journal for Quality in Health Care, 29-36.

Chakrabarti, A., Adams, C. E., Rathbone, J., Wright, J., Xia, J., Wong, W. et al. (2007). Schizophrenia trials in China: a survey. *Acta Psychiatr.Scand., 116,* 6-9.

Chong, S.-A., Ravichandran, N., Poon, L.-Y., Soo, K.-L., & Verma, S. (2006). Reducing polypharmacy through the introduction of a treatment algorithm: Use of a treatment algorithm on the impact on polypharmacy. *Annals of the Academy of Medicine Singapore,* 457-460.

Drake, R. E., Bond, G. R., & Essock, S. M. (2009). Implementing evidence-based practices for people with schizophrenia. *Schizophrenia Bulletin, 35,* 704-713.

Janssen, B., Ludwig, S., Eustermann, H., Menke, R., Haerter, M., Berger, M. et al. (2010). Improving outpatient treatment in schizophrenia: Effects of computerized guideline implementation-results of a multicenter-study within the German Research Network on Schizophrenia. *European Archives of Psychiatry and Clinical Neuroscience.260 (1) ()(pp 51-57), 2010.Date of Publication: February 2010.,* 51-57.

Karamustafalioglu, O. (2010). Major depressive disorder, mental health care, and the use of guidelines in the Middle East. *J Clin Psychiatry, 71 Suppl E1,* e07.

Koesters, M., Zhang, Y., Ma, Y., Weinmann, S., Becker, T., & Jin, W. (2010). What can we learn from Chinese randomised controlled trials? A systematic review and meta-analysis of Chinese venlafaxine studies.
Ref Type: Unpublished Work

Linden, M. (2007). A randomized controlled clinical trial comparing "guideline exposed" and "guideline naive" physicians in respect to dosage selection and treatment outcome with doxepin in depressive disorders. *Pharmacopsychiatry, 40,* 77-81.

Linden, M. (2008). Impact of the WHO depression guideline on patient care by psychiatrists: a randomized controlled trial. *European psychiatry : the journal of the Association of European Psychiatrists, 23,* 403-408.

Michaud, L., Voellinger, R., Burnand, B., & Stiefel, F. (2006). Major depressive disorders in the general hospital: How to implement guidelines. *Journal of Psychosomatic Research,* 455-459.

Morriss, R. (2008). Implementing clinical guidelines for bipolar disorder. *Psychology and Psychotherapy: Theory, Research and Practice,* 437-458.

Weinmann, S., Koesters, M., & Becker, T. (2007). Effects of implementation of psychiatric guidelines on provider performance and patient outcome: Systematic review. *Acta Psychiatrica Scandinavica,* 420-433.

Weinmann, S., Hoerger, S., Erath, M., Kilian, R., Gaebel, W., & Becker, T. (2008). Implementation of a schizophrenia practice guideline: Clinical results. *Journal of Clinical Psychiatry, Vol.69,* 1299-1306.

Zhang, M. (2010). Major depressive disorder treatment guidelines in China. *J Clin Psychiatry, 71 Suppl E1,* e06.

Appendix – Ovid Search Strategy

1. Guideline adherence

2. exp Guideline adherence/

3. practice guideline

4. exp practice guideline/

5. guideline$

6. professional standard

7. algorithm

8. or/1-7

9. implement$

10. strateg$

11. Evalua$

12. Assess$

13. validity

14. effect$

15. disseminat$

16. distribu$

17. issu$

18. impact

19. compar$

20. introduc$

21. reminder$

22. prompt$

23. feedback$

24. or/9-23

25. randomized controlled trial.pt.

26. controlled clinical trial.pt.

27. exp Randomized Controlled Trials/

28. random allocation.ab,hw,ot,sh,ti.

29. exp Random Allocation/

30. random$.ti.

31. exp Double-Blind Method/

32. exp Single-Blind Method/

33. ((singl$ or doubl$ or tripl$ or trebl$) and (blind$ or mask$ or dummy$)).ab,hw,ot,sh,ti.

34. (random$ and (trial or study)).ab,hw,ot,sh,ti.

35. exp intervention studies/

36. pre test or pretest or post test or posttest

37. comparative study

38. exp clinical trials/

39. clinical trial

40. (clin$ adj25 trial$).ti,ab.

41. exp evaluation studies/

42. follow-up studies.sh.

43. prospective studies.sh.

44. intervention?.tw.

45. or/25-44

46. exp mental health/

47. exp psychiatry/

48. psych$

49. mental health

50. mental ill$

51. psychiatry

52. exp psychology/

53. psychology

54. exp depression/

55. exp schizophrenia/

56. exp Bipolar Disorder/

57. exp Depressive Disorder/ or exp Psychotic Disorders/

58. or/46-57

59. 8 and 24 and 45 and 58