PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Identification of the delivery of cognitive behavioural therapy for psychosis (CBTp) using a cross sectional sample from Electronic Health Records and Open-Text Information in a Large UK based Mental Health Case Register
AUTHORS	Colling, Craig; Evans, Lauren; Broadbent, Matthew; Chandran, David; Craig, Thomas; kolliakou, Anna; Stewart, Robert; Garety, Philippa

VERSION 1 - REVIEW

REVIEWER	Richard Morriss University of Nottingham, UK I am a theme lead for NIHR MindTech and a work package lead for digital mental health interventions and primary care databases in a NIHR Biomedical Research Centre that examine care for people with mood disorders. I am NICE clinical expert.Therefore I might be seen as being in favour of the use of databases like CRIS (although I do not use this) and any research exploring issue related to exploring the use of NICE guidance in clinical practice.
REVIEW RETURNED	25-Jan-2017

GENERAL COMMENTS	The study reports a potentially important new utilisation of NHS records to obtain a more complete pictur3of the use of psychological treatments in routine NHS practice. I am not sure that I entirely agree that it it is the easiest approach to take but it certainly promises to be more complete and precise way of measuring the uptake of CBT. In fact the data showing that people from "black" ethnic groups in particular highlight important health inequalities that needs urgent action, and perhaps would not have been detected form small manual incomplete and selective audits that are currently performed. However, there are many important methodological omissions and questions that need to be addressed in relation to both the numerator and denominator. First of all what is the denominator here. Is this every person with a recorded diagnosis of schizophrenia or psychosis who had any contact with mental health services within the 12 month period? Does this include accident and emergency contacts, crisis teams, Improving Access to Psychological Treatments services, non-mental health services provided by the Trust, the national centres taking referrals from anywhere in the country , one off assessments or only service users accessing particular services. Did you include people with other comorbidities including those that might put the diagnosis of psychosis in doubt such as drug misuse, epilepsy, probable but uncertain diagnosis, people with shifting diagnosis e.g. bipolar disorder to schizophrenia

	etc.Overall what do you know about the completeness of recording of CRIS? Are there any services including those that might offer CBT that do not use CRIS? How do you ensure that multiple entries of the same person are counted as such rather than counted as different people and what about errors and completeness of data in relation to ethnicity and marital status. In relation to ethnicity, what does "black" cover? Where do South Asians fit and what about people of mixed race? Given that "black" is the biggest category was it really necessary to lump what might be very different ethnic groups under one ethnic category. Given the findings it is imperative to explain this and perhaps have an appendix giving more detail of the ethnic composition and use of CBT. Similarly how is marital status defined. Does this apply to people who are formally married or does it include civil partnerships, common law relationships or even people sharing the same living space? Was age related to a particular date in the 12 month period? In terms of the numerator, on the basis of the information supplied, I am not sure whether CBT for psychosis can be determined from CBT for other clinical problems. People may prefer other NICE recommended psychological treatments like family therapy. How could CBT be distinguished form other similar interventions such as coping strategies that might have a cognitive or behavioural element? The rates of delivery of CBTp are much higher than any previous estimate. SLAM is a major research and teaching centres is the measurement of CBT routine NHS activity delivered through the service or does this include research studies that are externally resourced and under different governance arrangements, and training for a diverse group of staff who may not otherwise deliver it in routine care? Is one session of CBTp an assessment session which fails to engage the participant or an assessment that the person is unsuitable for CBT. Could these be booster sessions? It is a pity that there is no assessment of completed courses of CBT. I am not sure that people receiving one or even 2 sessons of CBT are really what NICE had in mind. I am also not sure that people are expected to have continuous CBTp every year so what evidence is there that those who did not receive it had received CBT previously.? How many of the participants continuously receive CBT year after year? I would not necessarily expect data on all of this but there could be checks to see if the participants who took part in the audit in 2015 (the year first mentioned only in the discussion, not the method) also received CBT in 2013. These issues should be naswered where possible and acknowledged as limitations in the discussion. Another big issue to remedy (NAS2 and Haddock) which give vastly different results. The non-specialist reader is not provided any definition of these two crietia and cannot judge the usefulness of the data nor replicate the study unless the two are defined and the discussion pins down the importance of the findings that the two estimates of the delivery of CBT vary more than two-fold. Also if there are two sets of criteria is one applied to CBTp sessions given on one occasion or more and then two or more and the other only to the delivery of one session or more. Why is one set of criteria used for the further regression analysis and the other is not? How do these two sets of criteria relate to the more detailed audit by a senior clinician? Also why was the demographic and service delivery of CBTp assessed in the two different years (2013 and 2015) to check the robustness of the findings given the evident health inequalities

	that the presented analysis reveals. I do not understand why the terms recall and precision are used if they mean exactly the same as sensitivity and positive predictive value. I am not sure what figure 1 is supposed to show and that is not helped by having no title for the figure which should be self-explanatory. I am also not sure figure 2 relates to and so far have found no reference ot it in the rest of the text. It also does not have a title explaining what it shows. I ma also not sure why Table 2 has partially adjusted when unadjusted and fully adjusted would provide all the information that is required. In the Discussion, there is a statement that this method is easier and less time consuming. In the service I work in within my NHS Trust there is a simple proforma completed by anyone who delivers psychological treatment and this proforma is in the electronic notes. A data clerk can then determine how many people received CBT and what diagnosis is recorded. Isn't this simper than this method? There are however other advantages of this method in the paper so I think these need to be outlined more precisely. The methods they outline are far from simple and straightforward requiring people with special expertise that are not routinely employed in every NHS Trust. The richness of the data requires specific ethics and data protection checks. The discussion should really touch upon some of the issues in defining the numerator and denominator as outlined. Finally I don't think it is wise for the authors to publish these results without outlining what further steps are being taken to examine the health inequalities that their data show (not necessary in detail). I assume that some further investigation is going on to at least understand why people of black origin are so much less likely to receive CBTp than white or other people.

REVIEWER	Bernd Puschner Department of Psychiatry II, Ulm University, Germany
REVIEW RETURNED	01-Feb-2017

GENERAL COMMENTS	The authors describe a new method to measure whether Cognitive Behavioral Therapy for Psychosis (CBTp) is used by people with psychosis served by a large UK healthcare trust (South London and Maudsley NHS Foundation Trust, SLaM). The paper is well-written, but has some problems. 1. Abstract 1.1. “with high precision and recall”; It does not become clear what these terms mean (positive predictive value and sensitivity, explained way further down in the paper). Enough to report these indices in the results, but please generally avoid such uncommon technical terms. 1.2. “novel”; Apparently this method has been previously used for other research questions in many areas of medicine, apart from the very specific topic of provision of CBTp. Please omit. 1.3. “Receipt of CBTp was found to differ…“ In what way? Please

	report results. 1.4. I take issue with many of the conclusions: 1.4.1. It has not been demonstrated convincingly that this method is efficient and does save time. Obviously, it also involved a lot of manpower (staff effort of “human annotators” etc.), and to prove the point, time (per case) should have been compared among the three approaches (own, NAS21, and Haddock and colleagues2) which has not been done. 1.4.2. “is likely to be more accurate”; also this has not been proven as there was not direct comparison of the accuracy among the three approaches. High rates of positive predictive value and sensitivity are not enough. And the finding that rates of CBTp identified in this sample are higher than those found in comparable assessments1,2 also does not necessarily speak to the superiority of the method used in this paper (see 2.2). 1.5. Overall, the topic of the paper appears very specific to the UK. In which way are results of interest to an international audience? 2. Introduction 2.1. The introduction is well-written, but does not address readers outside the UK. What is the uptake of CBTp elsewhere? What is the international evidence, also from outside the UK? The authors seem to be aware of this shortcoming (“This paper is focused… of international concern and relevance.[11]”), but do not address it. 2.2. “The decision to focus initially on CBTp delivery instead of CBTp offer was a pragmatic one based on the perceived complexity and the resultant time required for each project.” This is a serious limitation which sheds doubt on the validity of findings because the two sources of evidence to which findings are compared applied a combined analysis of offer and uptake. Thus, the approach of the authors excludes those who were not offered CBTp or where no such offer was recorded. For a valid comparison, both offer and uptake should have been evaluated together. When taking a closer look at NAS21, one gets the impression (not 100% sure though) that uptake rates were restricted to those who had received an offer (“white diamonds” in Fig. 25 on p. 108). It has certainly been done that way by Haddock and colleagues2 (p. 163: “Of the 187 service users reviewed, 36 (19.3%) had evidence in… their clinical notes that they were verbally offered some form… of individual psychological therapy… Twenty four OF THESE 36 service users were then formally referred for individual psychological therapy in writing and of those, only 12 (6.4%) were specifically referred for therapy described as individual CBT.”). Obviously, leaving out offer from the analyses will produce higher uptake rates. 2.3. Please also report evidence about the method applied (NLP) which has obviously already been used in other areas. 3. Methods 3.1. “The initial step was to identify the delivery of CBT… with a diagnosis of psychosis (CBTp).”; “…with a diagnosis of psychosis the term CBTp is used from this point forward.” Then strictly speaking CBTp per se was not the basic unit of analysis, but CBT delivered to people with psychosis, which is not the same thing. 3.2. “… from free text fields within the BRC Case Register.” Is it mandatory for clinicians to enter CBT when it was provided? How sure can you be that CBT has been entered when provided? 3.3. Please explain in the text what is shown in figure 1. It appears that the information shown is merely conceptual (to explain the method and concepts), but not actually based on data. Please clarify. 3.4. “Once we were happy…”; Please do not use such colloquial and

	vague terms. Rather specify clear criteria (cf. 2.2). 3.5. "…classified as a psychological therapy in structured data drop down menu, …"; One gets the idea that now the very data source was used which has previously been criticized (p. 5: "there is concern that, as non-mandatory, it is incomplete."). Please clarify. 3.6. "The output… to be developed psychosis." I could not understand this sentence which is convoluted and ends oddly. Please correct. 3.7. Please do not repeat an expanded version of the research question in the methods section, but present your complete research questions at the end of the introduction. 4. Results 4.1. Sample descriptions miss a lot of important information (only have age and rough diagnoses for NAS2). Please add characteristics of study participants (e.g. demographic, clinical, social) as required by STROBE item 14. 4.2. "The SLaM return for the actual NAS2…" cf. 2.2. 5. Discussion 5.1. Please report evidence about NLP in more detail in the introduction, and use the information to interpret findings here (cf. 2.3). 5.2. "while also providing a much more comprehensive and accurate overview of the delivery of CBT across all cases". "This suggests that manual audits may result in under-reporting…" I do not consider these conclusions as warranted (cf. 2.2). 5.3. "Clearly there are also a large number of other variables…" Then why weren't these examined? This is no strength of this paper. 5.4. "The opportunity provided by employing methods shown here…" After learning about all the efforts needed to extract simple indicators of psychotherapy provision (and in my view not particularly valid ones) when reviewing this paper, wouldn't it be worthwhile to consider introducing a simple documentation system which records offer (yes, date) and number of CBTp sessions (date) per service user served by a trust? References 1 Royal College of Psychiatrists. Report of the second round of the National Audit of Schizophrenia (NAS2) 2014. http://www.rcpsych.ac.uk/workinpsychiatry/qualityimprovement/nationalclinicalaudits/nationalschizophreniaaudit/reports.aspx (accessed Feb 01, 2017). 2 Haddock G, Eisner E, Boone C, Davies G, Coogan C, … Barrowclough C. An investigation of the implementation of NICE-recommended CBT interventions for people with schizophrenia. Journal of Mental Health 2014;23:162–5. Taken together, the paper is well-written, but rather technical and of questionable relevance to readers outside the UK. Validity of findings appears doubtful. Reasons for increased uptake rates might be due to having used different methods than comparable studies (limiting to uptake of CBTp, ignoring offer). Revision is possible, but requires a lot of work, including reanalyses of the data.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Richard Morriss
Institution and Country: University of Nottingham, UK
Please state any competing interests or state 'None declared': I am a theme lead for NIHR MindTech and a work package lead for digital mental health interventions and primary care databases in a NIHR Biomedical Research Centre that examine care for people with mood disorders. I am NICE clinical expert. Therefore I might be seen as being in favour of the use of databases like CRIS (although I do not use this) and any research exploring issue related to exploring the use of NICE guidance in clinical practice.

Reviewer comment
The study reports a potentially important new utilisation of NHS records to obtain a more complete picture of the use of psychological treatments in routine NHS practice. I am not sure that I entirely agree that it is the easiest approach to take but it certainly promises to be more complete and precise way of measuring the uptake of CBT. In fact the data showing that people from "black" ethnic groups in particular highlight important health inequalities that needs urgent action, and perhaps would not have been detected form small manual incomplete and selective audits that are currently performed.

Author response
Thank you for highlighting the importance of this approach to identify health inequalities; we have therefore added this as strength of the paper within the discussion. We agree that this is not necessarily the easiest approach and within the discussion we emphasise that, once developed, this has the potential for being an efficient and effective method to provide a more comprehensive evaluation of CBTp which can be repeated at regular intervals.

Reviewer comment
However, there are many important methodological omissions and questions that need to be addressed in relation to both the numerator and denominator. First of all what is the denominator here. Is this every person with a recorded diagnosis of schizophrenia or psychosis who had any contact with mental health services within the 12 month period? Does this include accident and emergency contacts, crisis teams, Improving Access to Psychological Treatments services, non-mental health services provided by the Trust, the national centres taking referrals from anywhere in the country , one off assessments or only service users accessing particular services.

Author response
We apologise that we have not made this clear. We recruited using the inclusion criteria of two published audits as outlined within the Method section - every person with a recorded diagnosis of schizophrenia and with at least 12 months contact with community services were included irrespective of source. The published inclusion criteria do not include people with only accident and emergency contacts, crisis teams, Improving Access to Psychological Treatments services, non-mental health services provided by the Trust or the national centres taking referrals from anywhere in the country, in line with the published audit criteria.

Reviewer comment
Did you include people with other comorbidities including those that might put the diagnosis of psychosis in doubt such as drug misuse, epilepsy, probable but uncertain diagnosis, people with shifting diagnosis e.g. bipolar disorder to schizophrenia etc.

Author response

We apologise that this is not clear. We recruited using the criteria of two published audits as outlined within the Method section which did not specifically exclude these comorbidities.

Reviewer comment
Overall what do you know about the completeness of recording of CRIS?

Author response
We have added a statement in the results which indicate the completeness rates of the demographic variables used within the analysis in the Results section.

Reviewer comment
Are there any services including those that might offer CBT that do not use CRIS?

Author response
All services within SLAM are represented in CRIS, but the scope of this paper was community services (early intervention and promoting recovery services) which are focussed on SMI. This is explicit within the Method section but we have added to the abstract.

Reviewer comment
How do you ensure that multiple entries of the same person are counted as such rather than counted as different people?

Author response
Thank you for raising this. If a participant met the inclusion criteria across multiple teams within the same service type, to avoid double counting, the episodes were merged by selecting the earliest episode start date and latest end date for those episodes and presented as a single episode of care and we have added a statement to clarify this in the Methods section.

Reviewer comment
In relation to ethnicity, what does "black" cover? Where do South Asians fit and what about people of mixed race? Given that "black" is the biggest category was it really necessary to lump what might be very different ethnic groups under one ethnic category. Given the findings it is imperative to explain this and perhaps have an appendix giving more detail of the ethnic composition and use of CBT. Similarly how is marital status defined? Does this apply to people who are formally married or does it include civil partnerships, common law relationships or even people sharing the same living space? Was age related to a particular date in the 12 month period?

Author response
Thank you for your comment. We use the standard “NHS ethnicity 16 definitions” but did collapse them for the analysis. The Black ethnic group includes Black African (N), Black Caribbean (M) , any other Black background (P) and South Asians were grouped within the Other group. We have added a section within the Methods outlining in more detail the groupings for Ethnicity and Marital Status. Age was defined for this analysis at the point the person was accepted by the team and we have now added a table within the Results section which outlines the CBT % split by the NHS ethnicity 16 groupings.

Reviewer comment

In terms of the numerator, on the basis of the information supplied, I am not sure whether CBT for psychosis can be determined from CBT for other clinical problems.

Author response
The denominator was defined based on criteria from the published audits – i.e. patients with a psychosis diagnosis. In terms of the development of the NLP application we tested the performance using PPV and sensitivity initially for all clinical problems and then independently for patients with a psychosis diagnosis. These tests proved successful. We aimed to detect CBT provided for people with a diagnosis of psychosis and have called this CBTp, as is common practice. We consider this is what was done for the published audits .

Reviewer comment
People may prefer other NICE recommended psychological treatments like family therapy.

Author response
Thank you for your comment. This is part of a wider stream of work, focussing on other therapy types and we currently developing an application for family therapy with the aim to provide a more complete picture of psychological therapy provision as outlined within the discussion.

Reviewer comment
How could CBT be distinguished from other similar interventions such as coping strategies that might have a cognitive or behavioural element?

Author response
Thank you for your comment. This is not a simple question to answer, as CBT is comprised of a number of cognitive and behavioural elements. As noted within the limitations, the tool does not attempt to provide an assessment of CBT quality but simply the detection of CBT delivery for people with psychosis (CBTp) as described by therapist recording.

Reviewer comment
The rates of delivery of CBTp are much higher than any previous estimate.

Author response
The rates are broadly comparable with the NAS2 return when we used two or more sessions of CBT to indicate delivery (26.4% using our method compared with 20% with manual audit return) but vary by the number of sessions and timeframe. The method developed here is likely to have both better retrieval rates and differences in level of therapy provision in comparison with the findings of Haddock et al.

Reviewer comment
SLAM is a major research and teaching centres is the measurement of CBT routine NHS activity delivered through the service or does this include research studies that are externally resourced and under different governance arrangements, and training for a diverse group of staff who may not otherwise deliver it in routine care?

Author response
The sample presented here is reflective of the local service provision for the community i.e. Early intervention and Promoting recovery services within SLAM for local people , but SLAM local services

may and do on occasion benefit from research funded clinical activity. We have added a comment about this within the discussion including points relating to generalisability.

Reviewer comment
Is one session of CBTp an assessment session which fails to engage the participant or an assessment that the person is unsuitable for CBT. Could these be booster sessions?

Author response
The assessment for CBT is not viewed here as CBT delivery as it could be part of an assessment of suitability for CBTp but a CBTp assessment would be classified as the start of treatment and booster sessions could be included.

Reviewer comment
It is a pity that there is no assessment of completed courses of CBT.

Author response
Thank you for raising this. This is indeed a planned piece of work included in the next steps, but this requires a dedicated analysis and is beyond the scope of this particular paper.

Reviewer comment
I am not sure that people receiving one or even 2 sessions of CBT are really what NICE had in mind. I am also not sure that people are expected to have continuous CBTp every year so what evidence is there that those who did not receive it had received CBT previously? How many of the participants continuously receive CBT year after year? I would not necessarily expect data on all of this but there could be checks to see if the participants who took part in the audit in 2015 (the year first mentioned only in the discussion, not the method) also received CBT in 2013. These issues should be answered where possible and acknowledged as limitations in the discussion.

Author response
Thank you for these comments. To investigate the proportion of participants that received CBT 'year on year', we reviewed the data as suggested by the reviewer to "check to see if the participants who took part in the audit in 2015 also received CBT in 2013". This check found that 13.8% (53/385) of the participants who received CBTp in 2015 had also received CBTp in 2013 which is now reported in the results section. In addition Figure 1 shows the percentage of patients receiving CBTp delivery per year using the NAS2 inclusion criteria. We have added the rationale for re-sampling in 2015 into the method section.

Reviewer comment
Another big issue to remedy (NAS2 and Haddock) which give vastly different results. The non-specialist reader is not provided any definition of these two criteria and cannot judge the usefulness of the data nor replicate the study unless the two are defined and the discussion pins down the importance of the findings that the two estimates of the delivery of CBT vary more than two-fold.

Author response
Thank you for your comment. The audit definitions are specified for both audits within the 'participants' section of the methods and we have additionally flagged up the differences in the inclusion criteria adopted by these two published audits at the beginning of the discussion.

Reviewer comment
Also if there are two sets of criteria is one applied to CBTp sessions given on one occasion or more and then two or more and the other only to the delivery of one session or more.

Author response
The background outlined one limitation of NAS2, i.e. the lack of guidance, so we interpreted this in two ways and presented the data as specified in the method section – 'taken up' was interpreted as either at least one session of CBTp or at least 2 sessions of CBTp. In order to apply the Haddock inclusion criteria, we defined this as one or more session of CBTp because Haddock et al report the range of CBTp interventions in their data set as between 1-18.

Reviewer comment
Why is one set of criteria used for the further regression analysis and the other is not?

Author response
We selected the largest sample to test the demographic predictors of CBTp. This has now been added as a point of clarification to the Method section.

Reviewer comment
How do these two sets of criteria relate to the more detailed audit by a senior clinician?

Author response
We apologise that the term audit here was misleading – this was part of the development phase of the application and we have amended this term from "audit" to "assessment" within the Methods section. The aim was to independently assess the quality of the structured drop down box recording.

Reviewer comment
Also why was the demographic and service delivery of CBTp assessed in the two different years (2013 and 2015) to check the robustness of the findings given the evident health inequalities that the presented analysis reveals?

Author response
Thank you for raising this matter. Yes this was the reason, as the data presented here represented the first time such inequalities were identified and the rationale is now included in the method section.

Reviewer comment
I do not understand why the terms recall and precision are used if they mean exactly the same as sensitivity and positive predictive value.

Author response
Thank you for your comment. The terms precision and recall are standard terminology for Information retrieval tasks in natural language processing; however, I agree that they are unlikely to be sufficiently informative for a potential readership and I have amended to positive predictive value and sensitivity respectively.

Reviewer comment
I am not sure what figure 1 is supposed to show and that is not helped by having no title for the figure which should be self-explanatory.

Author response
We apologise, it appears the figure title was lost in the process of uploading for submission and the purpose was to explain the relationship between PPV and sensitivity. In retrospect this was unnecessary and we have removed the image.

Reviewer comment
I am also not sure figure 2 relates to and so far have found no reference to it in the rest of the text. It also does not have a title explaining what it shows.

Author response
We apologise, it appears the figure title was lost in the process of uploading for submission. We have amended the image to include the title. Fig 2 (now Figure 1) is referenced within the Results section.

Reviewer comment
I am also not sure why Table 2 has partially adjusted when unadjusted and fully adjusted would provide all the information that is required.

Author response
Thank you for your comment. As there is a relationship between age and service (Early intervention services are by definition for a younger patient group) we included the partially adjusted model which excludes service as a predictor, in order to check whether the increased likelihood of younger people being more likely to receive CBT is still significant and I have added clarification on this within the Methods section.

Reviewer comment
In the Discussion, there is a statement that this method is easier and less time consuming. In the service I work in within my NHS Trust there is a simple proforma completed by anyone who delivers psychological treatment and this proforma is in the electronic notes. A data clerk can then determine how many people received CBT and what diagnosis is recorded. Isn't this simper than this method?
Author response
We agree and we did not claim that this method is ‘easier’; however, once developed, it does remove the need for audits involving manual data extraction and provides the scope to generate very rapidly large and comprehensive samples as demonstrated here, identifying a sample of 2,579 people within between 01/07/2012 and 01/07/2013 time frame, using particular inclusion criteria. In the limitations section we have discussed more fully about NLP and other methods, including the use of a written proforma.

Reviewer comment
There are however other advantages of this method in the paper so I think these need to be outlined more precisely. The methods they outline are far from simple and straightforward requiring people with special expertise that are not routinely employed in every NHS Trust. The richness of the data requires specific ethics and data protection checks.

Author response

We agree there are advantages once the tool has been developed which include enabling regular access and routine reporting, scope to link to all the other available data in the record and to provide a more comprehensive cohort. Furthermore the ability to do this at one site allows the potential for multi-site audits to be implemented at relatively little cost, resulting in more robust data for service improvement.

Reviewer comment
The discussion should really touch upon some of the issues in defining the numerator and denominator as outlined.

Author response
Thank you for your comment. As previously stated we have added a statement at the start of the discussion which should clarify our use of the differing inclusion criteria drawn from the published audits.

Reviewer comment
Finally I don't think it is wise for the authors to publish these results without outlining what further steps are being taken to examine the health inequalities that their data show (not necessary in detail). I assume that some further investigation is going on to at least understand why people of black origin are so much less likely to receive CBTp than white or other people.

Author response
We agree, and have amended in the discussion the text to outline the steps the service has taken to understand and address the identified health inequalities which include regular reporting and close monitoring, and an action plan.

Reviewer: 2
Reviewer Name: Bernd Puschner
Institution and Country: Department of Psychiatry II, Ulm University, Germany
Please state any competing interests or state ‘None declared’: None declared.
Please leave your comments for the authors below

The authors describe a new method to measure whether Cognitive Behavioral Therapy for Psychosis (CBTp) is used by people with psychosis served by a large UK healthcare trust (South London and Maudsley NHS Foundation Trust, SLaM). The paper is well-written, but has some problems.

1. Abstract

Reviewer comment
1.1. “with high precision and recall”; It does not become clear what these terms mean (positive predictive value and sensitivity, explained way further down in the paper). Enough to report these indices in the results, but please generally avoid such uncommon technical terms.

Author response
As indicated in the response to Reviewer 1 we have amended this nomenclature throughout the manuscript.

Reviewer comment

1.2. “novel”; Apparently this method has been previously used for other research questions in many areas of medicine, apart from the very specific topic of provision of CBTp. Please omit.

Author response
We have removed the term “novel” as requested.

Reviewer comment
1.3. “Receipt of CBTp was found to differ…“ In what way? Please report results.

Author response
We have added the significance levels for the fully adjusted model where they differed significantly.

Reviewer comment
1.4.1. It has not been demonstrated convincingly that this method is efficient and does save time. Obviously, it also involved a lot of manpower (staff effort of “human annotators” etc.), and to prove the point, time (per case) should have been compared among the three approaches (own, NAS21, and Haddock and colleagues2) which has not been done.

Author response
Thank you for your comment. We have amended the wording to include “once developed” this approach saves time when compared with a manual case note audit.

Reviewer comment
1.4.2. “is likely to be more accurate”; also this has not been proven as there was not direct comparison of the accuracy among the three approaches. High rates of positive predictive value and sensitivity are not enough. And the finding that rates of CBTp identified in this sample are higher than those found in comparable assessments1,2 also does not necessarily speak to the superiority of the method used in this paper (see 2.2).

Author response
To obtain a direct comparison we have further analysed the performance of the method used here against the actual participants within the SLAM return of the NAS2 (n 100) fully described in our response to point 2.2 below. In addition, the method described here enables the regular production of reports using large datasets for a variety of uses compared with small samples usually obtained within manual case note audits.

Reviewer comment
1.5. Overall, the topic of the paper appears very specific to the UK. In which way are results of interest to an international audience?

Author response
Thanks for your comment. The provision of evidence based psychological therapy is of international concern. CBTp is recommended internationally in national guidelines in many countries particularly in Europe and North America and this has been further emphasised in the introduction including reference to policy recommendations.

2. Introduction

Reviewer comment
2.1. The introduction is well-written, but does not address readers outside the UK. What is the uptake of CBTp elsewhere? What is the international evidence, also from outside the UK? The authors seem to be aware of this shortcoming ("This paper is focused… of international concern and relevance.[11]"), but do not address it.

Author response
Thank you for your comment. There are guidelines many countries e.g. Australia and New Zealand, Canada, Spain and USA which recommend CBT for psychosis as an evidence based treatment which have been included as references in the manuscript. However, to our knowledge, despite the acknowledged implementation problem, there are few if any non UK studies as yet reporting on uptake.

Reviewer comment
2.2. "The decision to focus initially on CBTp delivery instead of CBTp offer was a pragmatic one based on the perceived complexity and the resultant time required for each project." This is a serious limitation which sheds doubt on the validity of findings because the two sources of evidence to which findings are compared applied a combined analysis of offer and uptake. Thus, the approach of the authors excludes those who were not offered CBTp or where no such offer was recorded. For a valid comparison, both offer and uptake should have been evaluated together. When taking a closer look at NAS21, one gets the impression (not 100% sure though) that uptake rates were restricted to those who had received an offer ("white diamonds" in Fig. 25 on p. 108). It has certainly been done that way by Haddock and colleagues2 (p. 163: "Of the 187 service users reviewed, 36 (19.3%) had evidence in… their clinical notes that they were verbally offered some form… of individual psychological therapy… Twenty four OF THESE 36 service users were then formally referred for individual psychological therapy in writing and of those, only 12 (6.4%) were specifically referred for therapy described as individual CBT."). Obviously, leaving out offer from the analyses will produce higher uptake rates.

Author response
We apologise and the reviewer raises an important point. We have further reviewed the Haddock method and results and we previously reported CBT delivery of 5.3% based on the 187 participants reviewed, of whom 13 had evidence of receiving an offer of CBT, and of whom 10 actually received CBT. As the reviewer correctly points out, CBT delivery is nested within CBT offer. Haddock additionally reports that 2 participants received CBT without any evidence of an offer of CBT within their manual case review which demonstrates that the offer and delivery of CBT were identified independently. We have therefore amended the manuscript to report the CBT delivery of 6.4 % based on the 187 participants reviewed and the 12 participants that received CBT as these data do not rely on having an offer of CBT.

As a further validity check of the automated method described here we anonymously examined the participants of the actual SLAM return for the national NAS2 audit (n = 100) which identified 20% participants as receiving CBT. Using the automated method in CRIS we found that 22% of the original participants received CBT which is similar to the manual case note audit and independent of the offer of CBT. This manual audit was carefully conducted, time consuming for clinicians, audit staff and managers but provides a much smaller sample when compared with the 2,308 participants when using the automated method and the same NAS2 inclusion criteria but not restricting the sample to 100 participants. We have updated the methods and results to reflect this validation check.

Reviewer comment
2.3. Please also report evidence about the method applied (NLP) which has obviously already been used in other areas.

Author response
NLP has been used for some time including a range of applications in different disease areas and we added a paragraph relating to this in the discussion.

3. Methods

Reviewer comment
3.1. “The initial step was to identify the delivery of CBT… with a diagnosis of psychosis (CBTp).”; “…with a diagnosis of psychosis the term CBTp is used from this point forward.” Then strictly speaking CBTp per se was not the basic unit of analysis, but CBT delivered to people with psychosis, which is not the same thing.

Author response
We appreciate the question. There is in fact debate amongst clinicians as to whether CBTp refers to an exclusive focus of CBT on symptoms of psychosis, or whether it encompasses all CBT provided for people with psychotic disorders. We adopted the latter approach and definition, and consider this most closely matches the approach of the two published audits, and have sought to make this clear in the Method section.

Reviewer comment
3.2. “… from free text fields within the BRC Case Register.” Is it mandatory for clinicians to enter CBT when it was provided? How sure can you be that CBT has been entered when provided?

Author response
The introduction highlights that the CBT structured drop down box is incomplete as a non-mandatory item. The majority of the clinical information is contained within free text of the case note, in which it is procedurally and medico-legally mandatory to record and describe clinical contact. One of the advantages of our approach was the opportunity to use and compare multiple sources to detect CBT delivery and we have highlighted within the results that the NLP application identified 21% additional service users who received CBTp but for whom this was not recorded in the dropdown box.

Reviewer comment
3.3. Please explain in the text what is shown in figure 1. It appears that the information shown is merely conceptual (to explain the method and concepts), but not actually based on data. Please clarify.

Author response
This was to explain the method and concepts. I have removed this from the manuscript.

Reviewer comment
3.4. “Once we were happy…”; Please do not use such colloquial and vague terms. Rather specify clear criteria (cf. 2.2).

Author response

I have amended this to the criterion applied, which was when the mean of the PPV and sensitivity were greater than 85%.

Reviewer comment
3.5. "…classified as a psychological therapy in structured data drop down menu, …"; One gets the idea that now the very data source was used which has previously been criticized (p. 5: "there is concern that, as non-mandatory, it is incomplete."). Please clarify.

Author response
Within the development process we investigated whether combining other available variables in addition to the NLP application output could improve performance. As the results indicate we excluded CBT sentences that commenced after the first 200 characters of the clinical document as a post processing rule to improve performance. I have updated the methods to reflect this.

Reviewer comment
3.6. "The output… to be developed psychosis." I could not understand this sentence which is convoluted and ends oddly. Please correct.

Author response
I have corrected this error.

Reviewer comment
3.7. Please do not repeat an expanded version of the research question in the methods section, but present your complete research questions at the end of the introduction.

Author response
The research question has been moved as requested.

4. Results

Reviewer comment
4.1. Sample descriptions miss a lot of important information (only have age and rough diagnoses for NAS2). Please add characteristics of study participants (e.g. demographic, clinical, social) as required by STROBE item 14.

Author response
We have updated the sample descriptions to include gender, ethnic origin, diagnostic groups and have added further detail for ethnic origin as requested by Reviewer 1.

Reviewer comment
4.2. "The SLaM return for the actual NAS2…" cf. 2.2.

Author response
Thank you for your comment. We think the response to cf.2.2. clarifies.

5. Discussion

Reviewer comment
5.1. Please report evidence about NLP in more detail in the introduction, and use the information to interpret findings here (cf. 2.3).

Author response
We have described where NLP applications have been used within health services and discussed issues relating to implementation within other NHS or services within other countries.

Reviewer comment
5.2. “while also providing a much more comprehensive and accurate overview of the delivery of CBT across all cases”. “This suggests that manual audits may result in under-reporting…” I do not consider these conclusions as warranted (cf. 2.2).

Author response
We have amended the Discussion section and have added text so that it is more evenly worded on the balance between manual audits and use of large-scale administrative data. We believe that there are advantages and disadvantages on both sides and that the two approaches should be considered as complementary rather than competing. On reflection, we feel that the original submission wasn’t sufficiently balanced on this issue, so we have sought to address this.

Reviewer comment
5.3. “Clearly there are also a large number of other variables…” Then why weren’t these examined? This is no strength of this paper.

Author response
We initially focussed on key demographic variables and this point relates to how the scope of the audits might be enhanced using this approach in comparison with manual case note audits which are labour intensive and limited in scope, as reflected in the discussion.

Reviewer comment
5.4. “The opportunity provided by employing methods shown here…” After learning about all the efforts needed to extract simple indicators of psychotherapy provision (and in my view not particularly valid ones) when reviewing this paper, wouldn’t it be worthwhile to consider introducing a simple documentation system which records offer (yes, date) and number of CBTp sessions (date) per service user served by a trust?

Author response
We appreciate the reviewer’s opinion and have tried to be more clear in the Discussion what this approach potentially adds. We emphasise that we do not see the two approaches as mutually exclusive. Manual documentation systems can be useful if there are local leads to monitor staff to enter data and use fields in a uniformed way; however, this itself can be labour intensive and is difficult to sustain over multiple areas of routine clinical practice. The NLP approach can be seen as complementary, as it has the capacity to handle clinician variability in practice and record-keeping, and (as mentioned above) to generate large and potentially multi-site samples for audits and research.

Reviewer comment
Taken together, the paper is well-written, but rather technical and of questionable relevance to

readers outside the UK. Validity of findings appears doubtful. Reasons for increased uptake rates might be due to having used different methods than comparable studies (limiting to uptake of CBTp, ignoring offer). Revision is possible, but requires a lot of work, including reanalyses of the data.

Author response
Thank you for your comments. We hope that our responses clarify the points raised. In terms of the relevance to readers outside the UK, we think that the policy and guidance recommending delivery of CBTp in many countries, and the acknowledged difficulty of implementation, highlight the importance of this method to assist in monitoring of the implementation.

VERSION 2 – REVIEW

REVIEWER	Richard Morriss Institute of Mental Health, University of Nottingham, United Kingdom
REVIEW RETURNED	28-Mar-2017

GENERAL COMMENTS	In terms of the limitations, the fact that the methods cannot distinguish who was offered CBT and then took up CBT, who was assessed and and then rejected for CBT after trying some sessions, who had a course of CBT and who had booster sessions, and whether they had any type of CBT or CBT for psychosis remains a significant limitation. I would be happy if this limitation was recognised in the limitations of the paper. The method can tell us who was delivered some CBT among people with psychosis in a large sample and highlighted an important inequality of provision showing the merit of such an approach; manual audits could not be done on this scale. I also think that more could be stated about the value of such informatics for monitoring the delivery of care for the international audience. The paper illustrates an important development in clinical informatics and the findings on inequalities shows added value that would be important for health systems internationally. I think the paper merits publication in this journal for these reasons.

REVIEWER	Bernd Puschner Department of Psychiatry II, Ulm University, Germany
REVIEW RETURNED	03-Apr-2017

GENERAL COMMENTS	My concerns have been adequately addressed. The paper has improved a lot.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Richard Morriss
Institution and Country: Institute of Mental Health, University of Nottingham, United Kingdom
Please state any competing interests or state ‘None declared: None

Please leave your comments for the authors below
Reviewer 1 comment
In terms of the limitations, the fact that the methods cannot distinguish who was offered CBT and then took up CBT, who was assessed and then rejected for CBT after trying some sessions, who had a

course of CBT and who had booster sessions, and whether they had any type of CBT or CBT for psychosis remains a significant limitation. I would be happy if this limitation was recognised in the limitations of the paper.
Author response
Thank you for your comment and we have updated the limitations as suggested.
Reviewer 1 comment
The method can tell us who was delivered some CBT among people with psychosis in a large sample and highlighted an important inequality of provision showing the merit of such an approach; manual audits could not be done on this scale. I also think that more could be stated about the value of such informatics for monitoring the delivery of care for the international audience. The paper illustrates an important development in clinical informatics and the findings on inequalities shows added value that would be important for health systems internationally. I think the paper merits publication in this journal for these reasons.

Author response
Thank you for your comment we have added the value of such informatics for monitoring care delivery for an international audience within the discussion.

Reviewer: 2
Reviewer Name: Bernd Puschner
Institution and Country: Department of Psychiatry II, Ulm University, Germany
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
My concerns have been adequately addressed. The paper has improved a lot.

VERSION 3 – REVIEW

REVIEWER	Richard Morriss University of Nottingham, UK
REVIEW RETURNED	10-Apr-2017

GENERAL COMMENTS	One minor typographical error I noticed on page 20, line 14 please remove the word "be" from between "is" and "of". Some of the paragraphs are rather long lasting most of a page. It would be easier to read if some of these were split up so one major point is equivalent to one paragraph.

VERSION 3 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Richard Morriss
Institution and Country: University of Nottingham, UK
Please state any competing interests or state 'None declared': None declared

Reviewer comment

One minor typographical error I noticed on page 20, line 14 please remove the word "be" from between "is" and "of".

Author response

I have amended the sentence as requested

Reviewer comment
Some of the paragraphs are rather long lasting most of a page. It would be easier to read if some of these were split up so one major point is equivalent to one paragraph.

Author response
I reviewed the document and split up some of the paragraphs as requested.