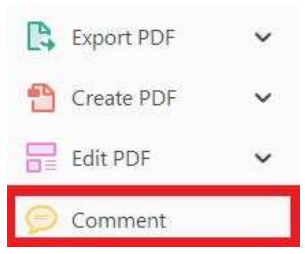
WILEY

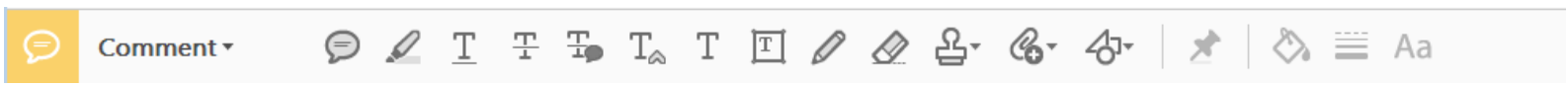
USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

Required software to e-Annotate PDFs: Adobe Acrobat Professional or Adobe Reader (version 11 or above). (Note that this document uses screenshots from Adobe Reader DC.)
The latest version of Acrobat Reader can be downloaded for free at: http://get.adobe.com/reader/

Once you have Acrobat Reader open on your computer, click on the Comment tab (right-hand panel or under the Tools menu).



This will open up a ribbon panel at the top of the document. Using a tool will place a comment in the right-hand panel. The tools you will use for annotating your proof are shown below:

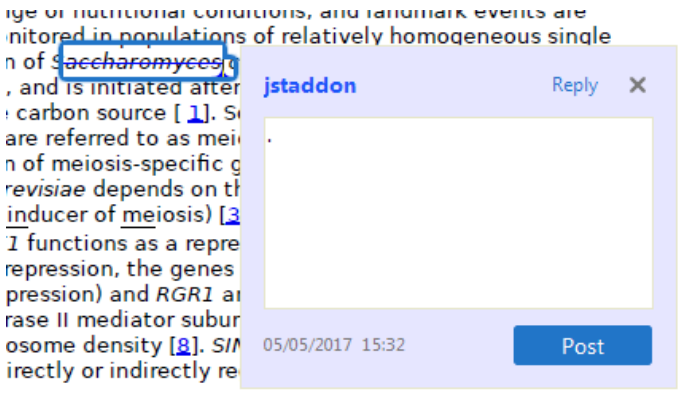


1. Replace (Ins) Tool – for replacing text.

Strikes a line through text and opens up a text box where replacement text can be entered.

How to use it:

- Highlight a word or sentence.
- Click on .
- Type the replacement text into the blue box that appears.

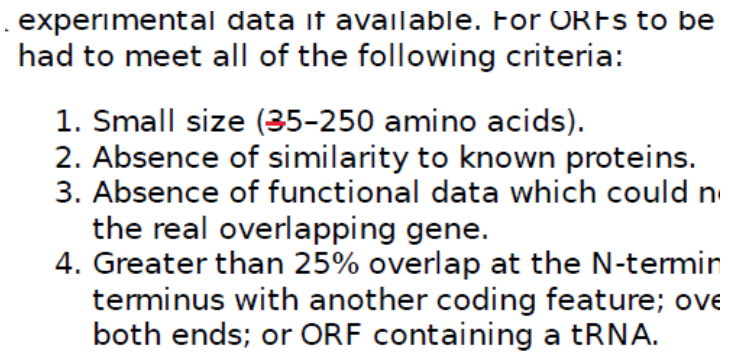


2. Strikethrough (Del) Tool – for deleting text.

Strikes a red line through text that is to be deleted.

How to use it:

- Highlight a word or sentence.
- Click on .
- The text will be struck out in red.


experimental data if available. For ORFs to be had to meet all of the following criteria:

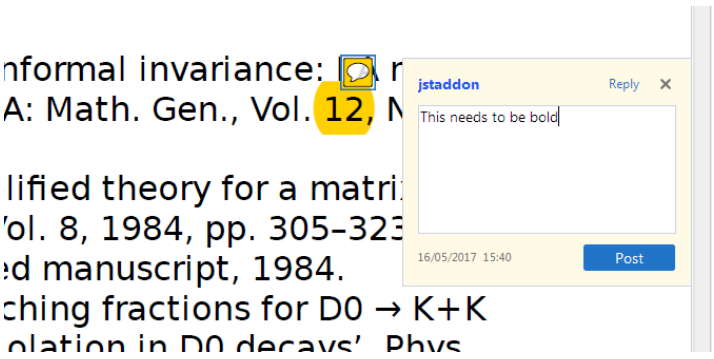
1. Small size (~~3~~5–250 amino acids).
2. Absence of similarity to known proteins.
3. Absence of functional data which could n the real overlapping gene.
4. Greater than 25% overlap at the N-termir terminus with another coding feature; ove both ends; or ORF containing a tRNA.

3. Commenting Tool – for highlighting a section to be changed to bold or italic or for general comments.

Use these 2 tools to highlight the text where a comment is then made.

How to use it:

- Click on .
- Click and drag over the text you need to highlight for the comment you will add.
- Click on .
- Click close to the text you just highlighted.
- Type any instructions regarding the text to be altered into the box that appears.

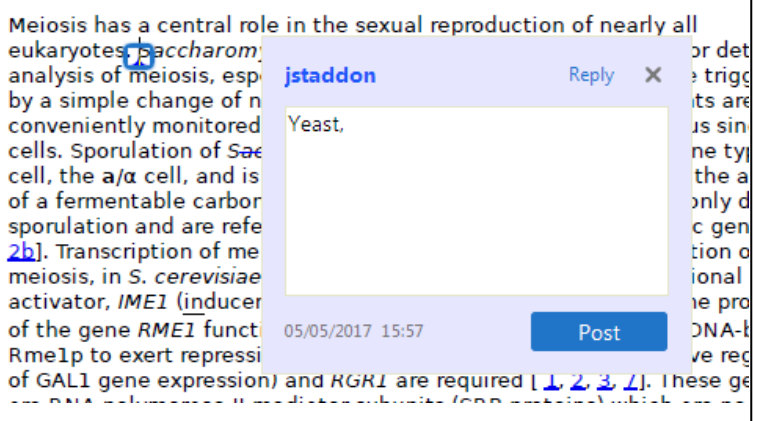


4. Insert Tool – for inserting missing text at specific points in the text.

Marks an insertion point in the text and opens up a text box where comments can be entered.

How to use it:

- Click on .
- Click at the point in the proof where the comment should be inserted.
- Type the comment into the box that appears.



WILEY

USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

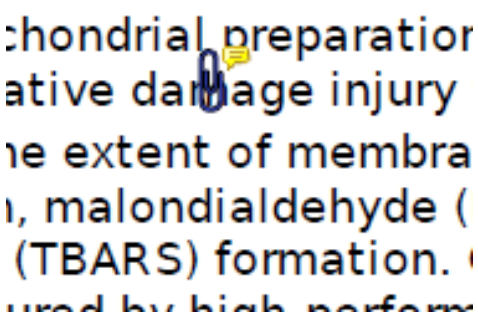
5. Attach File Tool – for inserting large amounts of text or replacement figures.

Inserts an icon linking to the attached file in the appropriate place in the text.

How to use it:

- Click on  .
- Click on the proof to where you'd like the attached file to be linked.
- Select the file to be attached from your computer or network.
- Select the colour and type of icon that will appear in the proof. Click OK.

  The attachment appears in the right-hand panel.

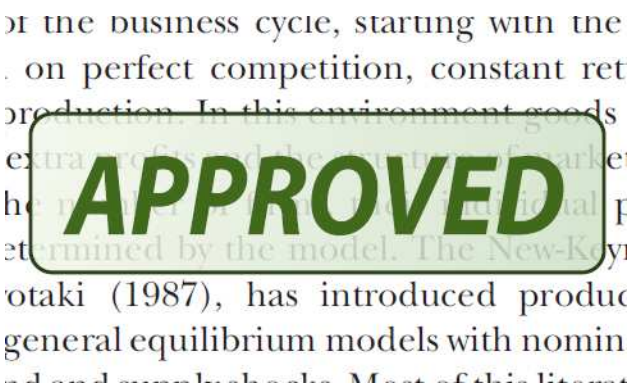


6. Add stamp Tool – for approving a proof if no corrections are required.

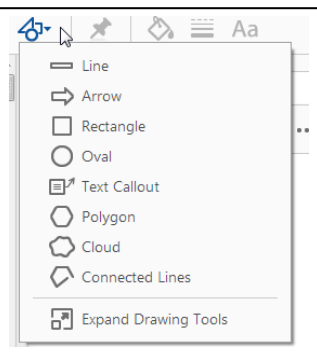
Inserts a selected stamp onto an appropriate place in the proof.

How to use it:

- Click on  .
- Select the stamp you want to use. (The Approved stamp is usually available directly in the menu that appears. Others are shown under *Dynamic*, *Sign Here*, *Standard Business*).
- Fill in any details and then click on the proof where you'd like the stamp to appear. (Where a proof is to be approved as it is, this would normally be on the first page).



Drawing tools available on comment ribbon

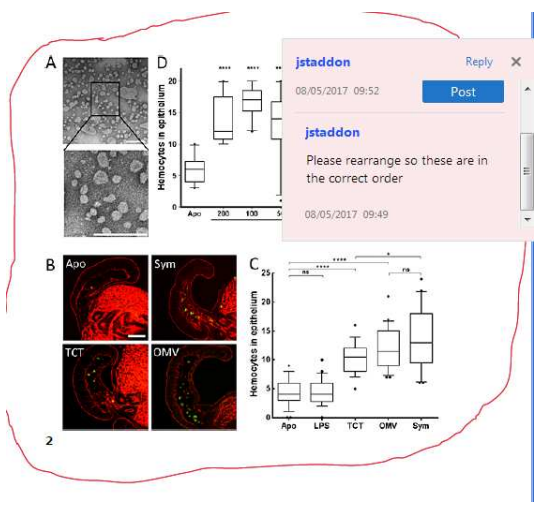


7. Drawing Markups Tools – for drawing shapes, lines, and freeform annotations on proofs and commenting on these marks.

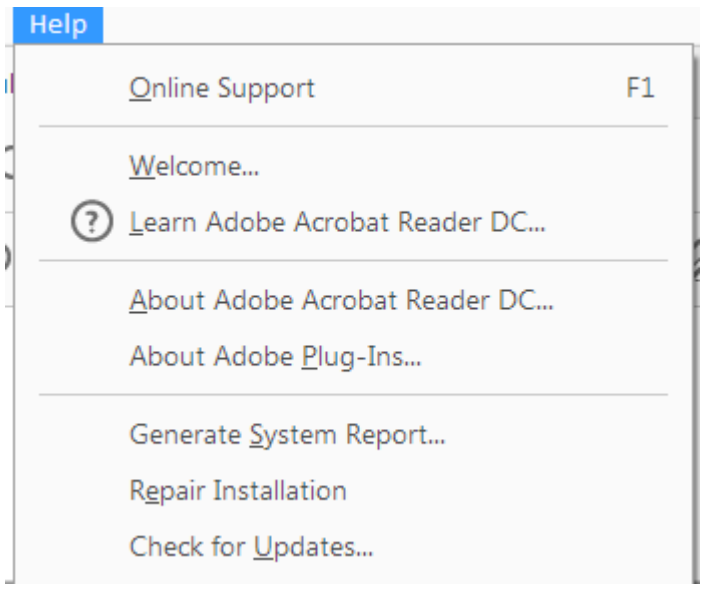
Allows shapes, lines, and freeform annotations to be drawn on proofs and for comments to be made on these marks.

How to use it:

- Click on one of the shapes in the Drawing Markups section.
- Click on the proof at the relevant point and draw the selected shape with the cursor.
- To add a comment to the drawn shape, right-click on shape and select *Open Pop-up Note.*
- Type any text in the red box that appears.



For further information on how to annotate proofs, click on the Help menu to reveal a list of further options:



WILEY

Author Query Form

Journal: Criminal Behaviour and Mental Health

Article: cbm_2104

Dear Author,

During the copyediting of your paper, the following queries arose. Please respond to these by annotating your proofs with the necessary changes/additions.

- If you intend to annotate your proof electronically, please refer to the E-annotation guidelines.
- If you intend to annotate your proof by means of hard-copy mark-up, please use the standard proofing marks. If manually writing corrections on your proof and returning it by fax, do not write too close to the edge of the paper. Please remember that illegible mark-ups may delay publication.

Whether you opt for hard-copy or electronic annotation of your proofs, we recommend that you provide additional clarification of answers to queries by entering your answers on the query sheet, in addition to the text mark-up.

Query No.	Query	Remark
Q1	AUTHOR: Please confirm that forenames/given names (blue) and surnames/family names (vermilion) have been identified correctly.	Yes
Q2	AUTHOR: Please verify that the linked ORCID identifiers are correct for each author.	Yes
Q3	AUTHOR: Please check that authors' affiliations are correct.	Yes
Q4	AUTHOR: Please check section headings if captured correctly according to its correct section levels.	Yes
Q5	AUTHOR: The citation “Quay, 1983” has been changed to “Quay, 1984” to match the author name/date in the reference list. Please check if the change is fine in this occurrence and modify the subsequent occurrences, if necessary.	Yes retain as Quay 1984
Q6	AUTHOR: Please check Tables 1 to 3 and their corresponding table notes if presented correctly.	See notes added
Q7	AUTHOR: Please provide a suitable caption for Tables 1 to 3.	See notes added
Q8	AUTHOR: “It was impossible to exclude cases where the behaviour was prevented by risk management, there was one completed suicide in the sample but many cases where suicidal behaviours were present but prevented by risk management such as being placed in an inpatient setting or residential care.” The meaning of this sentence is not clear. Please rewrite or confirm that the sentence is correct.	Include a full stop after "risk management" remove "such as" and change to "For example"

Query No.	Query	Remark
Q9	AUTHOR: Please provide a suitable caption for Figure 1.	See note ext to figure and movement of figure placement.
Q10	AUTHOR: Please confirm if the edit made in the sentence “Cooke and Michie (2013) and Hart, Michie, and Cooke (2007) have argued against the validity of actuarial measures in risk assessment of individual cases, valuing the primacy of clinical judgement within the framework of structured clinical or professional judgement tools.” conveys the intended meaning.	Yes but you can remove the added text "individual cases" and the meaning is still accurate.
Q11	AUTHOR: “Scurich & John, 2011” is cited in text but not provided in the reference list. Please provide details in the list or delete the citation from the text.	Please change to 2012
Q12	AUTHOR: “Andrews & Bonta, 1995” is cited in text but not provided in the reference list. Please provide details in the list or delete the citation from the text.	Please add this reference Andrews, D.A. , & Bonta, J. (1995). The Level of Service Inventory-Revised. Toronto, Canada: Multi-Health Systems
Q13	AUTHOR: Please check “ACKNOWLEDGEMENTS” and “CONFLICT OF INTEREST” if presented correctly.	One section changed
Q14	AUTHOR: "Andrews & Bonta, 2014" has not been cited in the text. Please indicate where it should be cited; or delete from the Reference List.	Yes delete please see section
Q15	AUTHOR: "Johnson et al, 2015" has not been cited in the text. Please indicate where it should be cited; or delete from the Reference List.	This was removed in past editorial changes. A sentence has been added to the text
Q16	AUTHOR: "Rice & Harris, 2005" has not been cited in the text. Please indicate where it should be cited; or delete from the Reference List.	Reference added to line 48 of page 4 of the article
Q17	AUTHOR: "Scurich & John, 2012" has not been cited in the text. Please indicate where it should be cited; or delete from the Reference List.	See question 11

Received: 28 March 2018 | Revised: 24 August 2018 | Accepted: 4 December 2018

DOI: 10.1002/cbm.2104

ORIGINAL ARTICLE

WILEY

Predictive validity and interrater reliability of the FACE-CARAS toolkit in a CAMHS setting

Stephen A. Evans[1] | David Young[2] | Paul A. Tiffin[3]

[1] Forensic Child and Adolescent Mental Health Service, NHS GGC Children and Young Peoples Specialist Services, West Glasgow Ambulatory Care Hospital, Glasgow, UK

[2] Department of Mathematics and Statistics, University of Strathclyde and NHS, Greater Glasgow and Clyde, Glasgow, UK

[3] Mental Health and Addiction Research Group, Department of Health Sciences, Faculty of Science, Alcuin Research Resource Centre, University of York, York, UK

Correspondence
Stephen A. Evans, Chartered Forensic Psychologist, Forensic Child and Adolescent Mental Health Service, NHS GGC Children and Young Peoples Specialist Services, West Glasgow Ambulatory Care Hospital, Dalnair Street, Glasgow G3 8SJ, UK.
Email: stephen.evans@ggc.scot.nhs.uk

Abstract

Background: The FACE-CARAS (Functional Analysis in Care Environments-Child and Adolescent Risk-Assessment Suite) toolkit has been developed to support practitioners in Child and Adolescent Mental Health Services (CAMHS) in performing a structured risk assessment. It covers a number of risk domains including violence, suicide, self-harm, experienced abuse, and exploitation. Interrater and internal reliability has already been established but not predictive validity.

Aims/Hypothesis: Our aim was to establish the predictive validity of the FACE-CARAS in a CAMHS population.

Methods: Records from 123 young people with FACE-CARAS ratings completed by clinicians were examined in a retrospective file review to extract data on a relevant list of adverse outcomes at three and at 6 months following the assessment. Although this was not a prospective longitudinal study, researchers were blind to the clinicians' ratings, allowing valid testing of predictive power. Cases were drawn from across generic and specialist CAMHS teams in approximately equal proportions. Data were analysed using receiver operator characteristic statistics.

Results: Areas under the curve values in five of the seven risk areas approached or were greater than 0.8 indicated that the FACE-CARAS profile score was a good potential predictor of risks of self-harm, suicidal behaviours, serious self-neglect, abuse or exploitation by others, and violence to others at both 3 and 6 months. It was weakly "predictive" of accidental self-harm and no better than chance at signalling physical ill health.

SPi | C B M | Journal Code | 2 1 0 4 | Article ID | Dispatch: 0.12.1 | No. of Pages: 10 | CE: Ma. Julianne Badoy Singco | ME:

Conclusions: Findings support the use of the "profile summary" section of the tool as likely to generate clinically useful risk predictions. We were concerned that clinical use of the scale did not conform to research standards and often left subscales incompletely rated; however, the fact that the tool nonetheless proved a good predictor of most key adversities under scrutiny may add weight to its value in clinical practice. Further work with the FACE-CARAS subscales is recommended.

KEYWORDS

CAMHS risk assessment, FACE-CARAS, risk assessment

1 | BACKGROUND

Q4

Risk assessment research in the criminal justice or forensic mental health system has focused mainly on risk of violence to others. The initial development of tools stemmed from offender classification systems (Blackburn, 1993, Quay, 1984) and focused on generating evidence that tools could classify groups of offenders for management Q5 and treatment regimes based on specific risk factors. Risk factors are developed from statistical regression, but when applied to young people, this creates particular problems as they have had less time to establish patterns of behaviour and because of the biological and social changes continuing to affect their development (Borum & Verhagen, 2006).

Tools for systematising assessment of risk of some adverse behaviours during adolescence have been developed, including risk of violence to others (Borum, Bartel, & Forth, 2006), of sexually harmful behaviour (Worling, Bookalam, & Litteljohn, 2012, and of general offending (Hoge & Andrews, 2002; McGrath & Thompson, 2012). There are fewer tools available to help simultaneously calculate risks *to* the young person, with attention only to more general risk factors (Posporelis et al., 2015). Suicide is rare in adolescence (Windfuhr et al., 2013), but it is estimated that 7–14% of adolescents have engaged in deliberately self-harming behaviours (Hawton & James, 2005), whereas accidental or coincidental self-harm, for example, from drug use or reckless driving, is also common.

Within risk assessment research, since 2010, there has been a focus on "Field Validity" (previously referred to as "ecological validity")—or how a risk assessment tool functions in a real world setting (see Perrault et al., 2017; Singh & Fazel, 2010). This focuses on how tools are used by clinicians and the predictions made function in real world practice. This approach has informed this study.

1.1 | U.K. service structures

Within U.K. NHS Child and Adolescent Mental Health Services (CAMHS), there is considerable political pressure on services to reduce time spent waiting for assessment and treatment. This has led to innovations such as the *Choice and Partnership Approach* (National CAMHS Support Service, 2009) that focus on streamlining and standardising the "patient's journey." This creates additional strain on professionals to categorise and predict risk to and from their patients with accuracy. U.K. NHS mental health services are organised into four tiers, reflecting complexity and level of intervention; those requiring management from mental health services are seen at Tier 3 generic regional CAMHS services. Services with additional specialist input, such as forensic mental health or learning disability psychiatry, are Tier 4.

The Functional Analysis in Care Environments (FACE) organisation developed the Child and Adolescent Risk Assessment Suite (CARAS) to support risk evaluation in CAMHS at the point the young person is first seen. The CARAS is innovative because it supports assessment of a wide range of risks not previously quantified. Although there have been prior attempts to analyse factors linked to adolescent vulnerability (Fischoff, Nightingale, & Ianotta, 2001), these were never brought together into a tool. The toolkit is now adopted by a number of NHS trusts in England, Scotland, Wales, and Northern Ireland. It covers the varied patterns of risk that confront the CAMHS worker and is a two-stage process. The worker must first review a checklist of 48 items and then decide on using one of nine component subscales: *self-harm*, *violence and aggression*, *vulnerability*, *risks associated with psychosis and violence*, *eating disorders*, *learning disability vulnerability*, *harmful sexual behaviour*, *specific management issues in open setting and in secure inpatient settings*. The assessor must rate each item on a scale of 0–3; scales vary in length but include around 20 individual factors. Finally, an overall rating of 0–4 is made for seven risks of *violence*, *suicide*, *deliberate self-harm*, *severe self-neglect*, *accidental self-harm*, *abuse/exploitation*, and *physical condition*. The anchor points for the scoring of this risk profile and implications for risk management are shown in Table 1. T1

The original content for the nine scales was drawn from published literature and findings from a series of focus groups (Daniel, Weir, & Tiffin, 2013). Subsequent work has established the interrater reliability of the scales in a field pilot (Tiffin, Kitchen, & Weir, 2015) and, in a more limited study, their interrater reliability (Evans & Oswald, 2017). These studies also provided evidence of content validity. The component scales can be reliably coded and scored, with minimal training, by practitioners. The predictive accuracy of the scales has not, however, previously been established. Our aim, therefore, was to evaluate the predictive accuracy of the summary risk profile scores.

2 | METHODS

2.1 | Ethics

The research protocol received a favourable ethical opinion from the West of Scotland Research Ethics Committee. The NHS Greater Glasgow and Clyde Research Department supported and sponsored the study.

2.2 | The sample

The sample was drawn from the NHS Greater Glasgow and Clyde Children and Young People's Specialist Services cover areas across Glasgow and the Clyde Valley. The area served has multiple issues of deprivation (Marryat, Thompson, Minnis, & Wilson, 2015), with resultant expectations of increased prevalence of adverse childhood

Q7 Q6 **TABLE 1** ...

The FACE risk scale	
0 = No apparent risk	No history/warning signs indicative of risk.
1 = Low apparent risk	No current indication of risk, but service user's history and/or warning signs indicate possible risk. Required precautions covered by standard care plan, that is, no special risk prevention measures or plan required.
2 = Significant risk	Service user's history and condition indicate the presence of risk, and this is considered to be a significant issue at present. Requires a contingency risk management plan.
3 = Serious risk	Substantial current risk. Circumstances are such that a risk management plan should be/has been drawn up *and* implemented.
4 = Serious and imminent risk	Service user's history and/or warning signs indicate the presence of risk, and this is considered imminent. Highest priority to be given to risk prevention.
n/k = Not known	Not currently enough information to determine a score.

Note. FACE: Functional Analysis in Care Environments.

experiences that, in turn, increase risk of psychiatric morbidity in children (Roberts, Donkin, & Marmot, 2016). All included patient records were from service Tiers 3 and 4.

Each team was asked to provide 10–15 cases that had been open for more than 6 months following risk assessment using the CARAS. No specific randomisation procedures were applied in selecting cases to be reviewed, but business support staff, who had no direct involvement in the cases, were used to provide case records. In some case notes, such as when the referral had been solely for a particular assessment, records were too limited to determine the presence or absence of behaviours after the assessment date, so these cases were removed from the final sample. This process resulted in inclusion of all Tier 4 cases with available records but an approximately 10% sample of Tier 3 cases with relevant records.

2.3 | Procedure for completing the FACE-CARAS by CAMHS clinicians

The procedure for completion of the FACE-CARAS has been discussed in the introduction. The clinician is expected to complete a checklist of risks after the initial assessment appointments, then select from a group of subscales covering patterns of risk related to eating disorders, physical aggression, self-harm and suicide, sexually harmful behaviour, vulnerability, and risks for inpatient populations and young people with learning disabilities. From this assessment, the clinician determines the "risk profile"—level of risk and risk management in the seven areas discussed.

2.4 | Procedure used by researchers to examine accuracy of prediction

The outcome data were collected from an exhaustive retrospective review of all of the case notes in existing mental health file records and was completed by the researcher. The presence or absence of risk-related behaviours was then recorded under the following categories—*violence*, *suicide*, *deliberate self-harm*, *severe self-neglect*, *accidental self-harm*, *abuse/exploitation*, and *physical condition*—as either present or absent. A record was made if the behaviour appeared in the first 3 months following assessment or in the second 3 months. If the behaviour was first found in the first 3 months, it was considered met in the second 3 months. This was then compared against the risk profile score ascribed by the clinician in the original FACE-CARAS Risk Profile Assessment.

2.5 | Defining outcome

Assessment of the value of a risk assessment tool depends comparing the extent to which a projected outcome is fulfilled by occurrence of a real outcome. The seven projected outcomes are defined by the scale scores, but the seven real outcomes have not previously been strictly defined by the authors of the scale. We therefore developed criteria for them (see Appendix S1). It was impossible to exclude cases where the behaviour was prevented by risk management, there was one completed suicide in the sample but many cases where suicidal behaviours were present but prevented by risk management ~~such as~~ being placed in an inpatient setting or residential care. This created some tautology for these factors: A high-risk assessment score was also likely to lead Q8 to inpatient care or use of child protection measures however risk management was invariably accompanied by risk-related behaviour.

2.6 | Statistical analyses

We used receiver operator characteristics to determine the significance and size of the effect which, effectively, which yields an "area under the curve (AUC), the size of which indicated the range of correct 'hits.'" An AUC of 1 would be perfect and of 0.5 not better than chance. It is generally accepte th t 0.9–1.0 represents excellent

prediction, 0.8–0.9 good prediction, 0.7–0.8 fair prediction, 0.6–0.7 weak prediction, and 0.5–0.6 no or little better than chance.

Two sets of scores are available from the tool produced by the risk profile giving a score of between 0 and 4 for each risk type as noted in the table above and scores from subscales used by the clinicians. The risk profiles were used universally, but few subscales were completed meaning that any analysis of the subscales was impossible. It had been intended to use a sum of scaled scores as a form of analysis of the predictive accuracy of subscales. This output research focuses solely on the risk profile scores, and no analysis is included of the subscales. Statistical support was provided through Strathclyde University.

3 | RESULTS

There were 123 young people for whom at least 6 months of adequate postassessment data were available. Mean age was 13.5 years (range 5–18 years), but 91 (74%) were 13–18. Eighty-nine of the sample were drawn from the eight regional NHS CAMHS (Tier 3) teams in the greater Glasgow area, and 34 were from the three specialist CAMHS teams (Tier 4). Two teams provided just eight cases and all the rest between the 10 and 15 requested.

3.1 | How the scale is used in practice

There are two parts to the CARAS process, risk profiles that were always completed and supplementary subscales related to specific risk patterns. Subscales of the CARAS had rarely been completed. The publishers provide no specific guidance on when a subscale should be used, so this is at the discretion of the clinician. The researchers, however, determined a score of "2" on the final risk profile was an indication that at least one full subscale should have been completed. When assessed against this standard, we found that Tier 3 teams had followed this in 25% of cases; among Tier 4 teams, relevant subscale completion was higher at 73%.

Final risk profile scores differed between the two tiers of service, with Tier 3 services recording one or more scores of or 2 or greater in 30 (34%) cases and Tier 4 in 17 (55%) cases. There was a difference in risk profile scores between Tier 3 and 4 teams, with patterns of risk profile scores significantly distinguishing between the tiers (f-test $p < 0.05$). A similar pattern was observed in relation to multiple risks, with three (3%) Tier 3 cases having more than one risk profile score at 2 or more and 8 (24%) of Tier 4 cases.

3.2 | Accuracy of prediction

Table 2 shows that five of the seven risk subscales had a fair to good prediction accuracy; risk of accidental harm was predicted with only modest accuracy. Only the risk related to physical condition scale offered a prediction that was no better than chance. The table also shows predictive values for outcomes at 3 and 6 months separately. T2

There is little difference between the AUC figures at 3 and 6 months, but if anything, a tendency for slightly improved prediction in some subscales and minor reduction in some other areas.

In the majority of assessments, the accuracy of prediction was heavily influenced by the levels of true negative predictions—where the clinician correctly predicted no risk. Figure 1 shows a distribution effect across the risk categories. The dense grouping at the score 0 denoting absence of risk of adverse events confirms clinicians' ability to categorise, essentially, safety correctly and that false positives were rare. F1

3.3 | Analysis of prediction by score point

Prediction generally functioned best at values greater than zero, and this is evident in the sensitivity and specificity scores shown in Table 3. This indicated a reluctance to give high scores, the most commonly assigned rating above 0 T3

TABLE 2

Risk type		*N* (123) Positive	Negative	AUC	95% Confidence interval	*Z* statistic	*p*
Deliberate self-harm	3 months	25 (20.33%)	98 (79.67%)	0.839	[0.761, 0.899]	7.902	<0.0001
	6 months	34 (27.64%)	89 (72.36%)	0.796	[0.714, 0.864]	6.751	<0.0001
Violence/Harm to others	3 months	36 (29.27%)	87 (70.73%)	0.851	[0.776, 0.909]	9.149	<0.0001
	6 months	38 (30.89%)	85 (69.11%)	0.884	[0.814, 0.935]	11.391	<0.0001
Suicidal behaviour	3 months	16 (13.01%)	107 (86.99%)	0.918	[0.855, 0.960]	14.981	<0.0001
	6 months	22 (17.89%)	101 (82.11%)	0.893	[0.824, 0.941]	14.320	<0.0001
Severe self-neglect	3 months	19 (15.45%)	104 (84.55%)	0.787	[0.704, 0.856]	4.873	<0.0001
	6 months	20 (16.26%)	103 (83.74%)	0.825	[0.746, 0.888]	6.127	<0.0001
Accidental self-harm	3 months	29 (23.58%)	94 (76.42%)	0.815	[0.735, 0.879]	7.478	<0.0001
	6 months	33 (26.83%)	90 (73.17%)	0.775	[0.691, 0.845]	5.953	<0.0001
Abuse/Exploitation	3 months	19 (15.45%)	104 (84.55%)	0.889	[0.820, 0.939]	9.682	<0.0001
	6 months	24 (19.51%)	99 (80.49%)	0.844	[0.767, 0.903]	7.369	<0.0001
Risk related to physical condition	3 months	3 (2.44%)	120 (97.56%)	0.790	[0.708, 0.858]	1.651	0.0988
	6 months	5 (4.07%)	118 (95.93%)	0.650	[0.559, 0.734]	1.160	0.2460

Note. AUC: area under the curve.

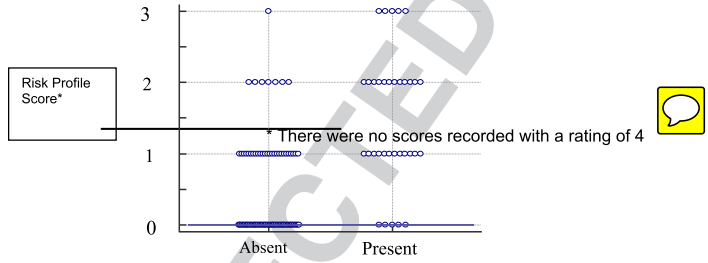


FIGURE 1 [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3

Risk category	Score > 0 Sens	Spec	PPV	NPV	Score > 1 Sens.	Spec	PPV	NPV	Score > 2 Sens	Spec.	PPV	NPV
Deliberate self-harm	85.29	61.80	46.0	91.7	50.00	91.01	68.0	82.7	14.71	98.88	83.3	75.2
Violence/Harm to others	86.84	84.71	71.7	93.5	55.26	96.47	87.5	82.8	15.76	98.82	85.7	72.4
Suicidal behaviour	100	63.37	37.3	100	59.09	92.08	92.08	91.02	18.18	100	100	84.9
Severe self-neglect	75.0	88.35	55.6	94.8	25	98.06	71.4	87.1	10	100	100	85.1
Accidental self-harm	81.82	64.44	45.8	90.6	50.0	97.98	60.9	81.0	33.33	100	100	76.9
Abuse/Exploitation	87.50	64.65	37.5	95.5	42.42	90.0	85.7	89.0	18.18	100	100	86.1
Risk related to physical condition	40	88.14	12.5	97.2	20.0	97.46	25.0	96.6	20.0	100	100	96.7

Note. Sens.: sensitivity; Spec.: specificity; PPV: positive predictive value; NPV: negative predictive value.

being "1." Although it may be a characteristic of the population in mainstream CAMHS that high risks are largely absent compared with specialist CAMHS, where risky behaviours appeared more frequently, it may be that clinicians are reluctant to produce high risk ratings on the profiles. This may be caused by workload stress and a wish to avoid the increase to the workload from higher risk ratings and related risk management.

4 | DISCUSSION

We found that CAMHS clinicians were able to use the CARAS to assess a range of seven major risks of harm for the young patients entering their services. Their resultant predictions fell into the good to excellent range in correctly identifying behaviours or events in relation to these young people at 3 months and at 6 months after the ratings. The only area of prediction which was apparently no better than chance was related to physical health, but this was an area where significant risks were exceptional, and a no risk rating was accurate.

Although AUC statistics compare favourably with other risk assessment tools, it should be remembered that many of the patients are being admitted to Tier 3 services for procedures such as diagnostic assessments, and high levels of risk would not normally be expected in these patients. A primary finding is that clinicians are able to differentiate these cases from cases with more obvious patterns of risk. This is a different risk assessment task than determining comparative levels of risk within populations such as released prisoners or comparative risk of neglect in children in care.

4.1 | What type of risk assessment is the FACE-CARAS?

Our findings show there is limited adherence to use of the full protocol for use of this scale, with the vast majority of assessments rarely using the subscale sections of the toolkit even when the screen score should trigger such action. This raises a question as to whether the measure is being used for simply recording unstructured professional judgement. Discussion around the use and validity of different risk assessment methods continues.

Cooke and Michie (2013) and Hart, Michie, and Cooke (2007) have argued against the validity of actuarial measures in risk assessment of individual cases, valuing the primacy of clinical judgement within the framework of structured clinical or professional judgement tools. This research has led to criticism (Mossman & Selke, 2007) and papers that have deconstructed the validity of the statistical measures used by Hart, Michie, and Cooke (Scurich & John, 2011). Other risk assessment researchers have argued that all risk assessment should be actuarial and that clinical judgement should be routinely distrusted (Harris, Rice, Quinsey, & Cormier, 2015). Both groups have provided empirical support and argument for their views and largely accept a "scholarly division of opinion" (Harris et al., 2015, p. 184).

It is important to emphasise that the clinicians in this study were forming their judgements using the FACE-CARAS toolkit but largely deviating from the scale protocols. The tool's structure and training indicate that the clinician should come to a judgement through a two-stage process, first completing a screening checklist of risk factors and then completing relevant subscales before coming to a final judgement. In the records assessed, it would appear that many were not following up on the subscales, but the checklist of common risk factors is enough to anchor the judgments anyway—and is what is achievable within workplace pressures.

The number of judgements being made at this point (48 separate risk factors rated at two time periods) would make it broadly similar to other older established risk assessments used for offender classification, namely, the Level of Service Inventory (Andrews & Bonta, 1995) and the Youth Level of Service Inventory (Hoge & Andrews, 2002), and places the FACE-CARAS alongside other proven structured professional judgement tools that are anchored by actuarial factors developed for classification of need. The protocol has been adapted by applied clinicians to meet their needs as service providers working under pressure but still retains accuracy.

4.2 | Study strengths and limitations

There are two significant limitations to this research. The sampling method was created to ensure a database of cases with adequate record keeping rather than using a randomised pattern of case selection. The eventual method of asking support staff rather than case managers to select the cases from the electronic databases prevented cherry

picking of cases with optimal outcomes, but we recognise randomisation would have improved the methodology. It should also be noted that our sample was of patients in the service for at least 6 months. Thus, our findings may not generalise to more short-term cases. Second, as discussed previously, there is a tautology in the structures used to determine the outcomes in suicide assessments and risk of abuse and exploitation. Consequently, the clinician can potentially reduce the chance of their own prediction being fulfilled, indeed has a duty to do so, by beginning processes to have the young person placed in hospital or residential care.

Our findings place the findings on accuracy of suicide risk assessment in conflict with recent research in this area that has found clinical judgement (Woodford et al., 2017) and risk scales (Chan et al., 2016) to be inaccurate predictors. The primary reason for this is these studies used completed suicide as the criteria for risk being present. Given the extensive legal powers that can be taken to prevent suicide or impose child protection, it is unclear how meaningful this is as a measure as it implies that risk management has broken down possibly indicating a failure in care rather than a characteristic of the patient. Whatever criteria is used will influence the assessment of predictive accuracy; however, if the criteria of competed suicide was to be used in this study, the predictive accuracy for this behaviour would be absent.

5 | CONCLUSIONS

We found that the FACE-CARAS shows good predictive accuracy over a range of risks of harm common in a child and adolescent mental health populations. A strength of our study was that it analysed the use of the scale in a real world setting, demonstrating its strengths even when abbreviated by clinicians in practice.

ACKNOWLEDGEMENTS

Q13

Thanks are given to Dr. Jennifer McDonald, Professor Helen Minnis, Lesley Dunabie, and Scott Wilson who supported the development of this project.

CONFLICTS OF INTEREST

There are no known conflicts of interests identified by the authors. The study has been sponsored by NHSGGC research department, and time for the research has been provided by NHSGGC Children and Young People's Specialist Services.

~~If this document is accepted for publication, it is intended to ask the test publisher to meet APC charges.~~

ORCID

Stephen A. Evans https://orcid.org/0000-0002-6695-0717

REFERENCES

~~Andrews, D. A., & Bonta, J. (2014). *The psychology of criminal conduct* (Fifth ed.). Cincinnati, OH: Anderson Publishing. https://doi.org/10.4324/9781315721279~~ Q14

Blackburn, R. (1993). *The psychology of criminal conduct: Theory, research and practice*. Chichester: John Wiley and Sons.

Borum, R., Bartel, P., & Forth, A. (2006). *Structured assessment of violence risk in youth: Professional manual*. Lutz. FL: Psychological Assessment Resources.

Borum, R., & Verhagen, D. (2006). *Assessing and managing violence risk in juveniles*. New York NY: The Guildford Press.

Chan, M. K. Y., Bhatti, H., Meader, N., Stockton, S., Evans, J., O'Connor, R. C., … Kendall, T. (2016). Predicting suicide following self-harm: Systematic review of risk factors and risk scales. *British Journal of Psychiatry*, *209*, 277–283. https://doi.org/10.1192/bjp.bp.115.170050

Cooke, D. J., & Michie, C. (2013). Violence risk assessment: From prediction of from what? To why? In C. Logan, & L. Johnstone (Eds.), *Managing clinical risk: A guide to effective practice* (pp. 3–25). Oxon: Routledge.

Daniel, M. R., Weir, S., & Tiffin, P. A. (2013). A novel CAMHS risk assessment system: Clinicians' views. *Journal of Forensic Practice*, *15*(3), 182–191. https://doi.org/10.1111/camh.12110

Evans, S. A., & Oswald, E. (2017) Inter-rater reliability of the FACE-CARAS Measure: Summary of available research and findings from a study using training materials. *Snapshot: Psychology across NHSGGC*, *3*. www.staffnet.ggc.scot.nhs.uk/Partnerships/MHP/Specialist%20Services/Psychology/Documents/Snapshot%20-%20Psychology%20Newsletter/Issue%203/Inter-rater%20reliability%20of%20the%20FACE%20CARAS%20scale%20Stephen%20Evans%20Eleanor%20Oswald%20December%202016.pdf

Fischoff, B., Nightingale, E. O., & Ianotta, J. G. (2001). *Adolescent risk and vulnerability: Concepts and measurement*. Washington, DC: National Academy Press.

Harris, G. T., Rice, M. E., Quinsey, V. L., & Cormier, C. A. (2015). *Violent offenders: Appraising and managing risk* (Third ed.). Washington DC: American Psychological Association. https://doi.org/10.1037/14572-000

Hart, S. D., Michie, C., & Cooke, D. J. (2007). Precision of actuarial risk assessment instruments. Evaluating the 'margins of error' of group v. individual predictions of violence. *British Journal of Psychiatry*, *190*(49), 60–65. https://doi.org/10.1192/bjp.190.5.s60

Hawton, K., & James, A. (2005). Suicide and deliberate self harm in young people. *British Medical Journal*, *330*, 891–894. https://doi.org/10.1136/bmj.330.7496.891

Hoge, R. D., & Andrews, D. A. (2002). *Youth level of service/case management inventory: User's manual*. Toronto, Canada: Multi Health Services.

Johnson, W., Clancy, T., & Bastian, P. (2015). Child abuse/neglect risk assessment under field practice conditions: Tests of external and temporal validity and comparison with heart disease prediction. *Children and Youth Services Review*, *56*, 76–85. https://doi.org/10.1016/j.childyouth.2015.06.013 Q15

Marryat, L., Thompson, L., Minnis, H., & Wilson, P. (2015). Exploring the social, emotional and behavioural development of preschool children: Is Glasgow different? *International Journal for Equity in Health*, *14*, 3. https://doi.org/10.1186/s12939-014-0129-8

McGrath, A., & Thompson, A. P. (2012). The relative predictive validity of the static and dynamic domain scores in risk-need assessment of juvenile offenders. *Criminal Justice and Behavior*, *39*(3), 250–263. https://doi.org/10.1177/0093854811431917

Mossman, D., & Selke, T. (2007). Avoiding errors about "margins of error". *British Journal of Psychiatry*, *191*, 561. https://doi.org/10.1192/bjp.191.6.561

National CAMHS Support Service (2009). *Evaluation of the choice and partnership approach in child and adolescent mental health services in England*. London: The Mental Health Foundation. https://www.mentalhealth.org.uk/sites/default/files/CAPA_PDF.pdf

Perrault, R., Vincent, G., Guy, L., Ben-Porath, Y. S., Edens, J. F., & Boccaccini, M. T. (2017). Are risk assessments racially biased?: Field study of the SAVRY and YLS/CMI in probation. *Psychological Assessment*, *29*(6), 664–678. https://doi.org/10.1037/pas0000445

Posporelis, S., Paspali, A., Takayanagi, Y., Sawa, A., Banerjea, P., & Kyriakopoulos, M. (2015). Demographic and clinical correlates of suicidality in adolescents attending a specialist community mental health service: A naturalistic study. *Journal of Mental Health*, *24*(4), 225–229. https://doi.org/10.3109/09638237.2015.1022249

Quay, H. C. (1984). *Managing adult inmates: Classification for housing and program assignment*. College Park, MD: American Correctional Association. ISBN-13: 978-0942974645

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d*, and *r*. *Law and Human Behavior*, *29*(5), 615–620. https://doi.org/10.1007/s10979-005-6832-7 Q16

Roberts, J., Donkin, A., & Marmot, M. (2016). Opportunities for reducing socioeconomic inequalities in the mental health of children and young people—Reducing adversity and increasing resilience. *Journal of Public Mental Health*, *15*(1), 4–18. https://search.proquest.com/docview/1776379779. https://doi.org/10.1108/JPMH-08-2015-0039

Scurich, N., & John, R. S. (2012). A Bayesian approach to the group versus individual prediction controversy in actuarial risk assessment. *Law and Human Behavior*, *36*(3), 237–246. https://doi.org/10.1037/h0093973 Q17

Singh, J. P., & Fazel, S. (2010). Forensic risk assessment: A metareview. *Criminal Justice and Behavior*, *37*(9), 965–988. ISSN: 00938548, https://doi.org/10.1177/0093854810374274

Tiffin, P. A., Kitchen, C. E. W., & Weir, S. (2015). Innovations in practice: Piloting a new child and adolescent risk assessment suite in the UK. *Child and Adolescent Mental Health*, *20*(4), 225–229. https://doi.org/10.1111/camh.12110

Windfuhr, K., While, D., Hunt, I. M., Shaw, J., Appleby, L., & Kapur, N. (2013). Suicide and accidental deaths in children and adolescents in England and Wales, 2001-2010. *Archives of Disease in Childhood*, *98*(12), 945–950. https://doi.org/10.1136/archdischild-2012-302539

Woodford, R., Spittal, M. J., Milner, A., McGill, K., Kapur, N., Pirkis, J., … Carter, G. (2017). Accuracy of clinician predictions of future self-harm: A systematic review and meta-analysis of predictive studies. *Suicide & Life-Threatening Behavior*, *3*, 1–18. https://doi.org/10.1111/sltb.12395

Worling, J. R., Bookalam, D., & Litteljohn, A. (2012). Prospective validity of the Estimate of Risk of Adolescent Sexual Offense Recidivism (ERASOR). *Sexual Abuse: A Journal of Research and Treatment*, *24*(3), 203–223. https://doi.org/10.1177/1079063211407080

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Evans SA, Young D, Tiffin PA. Predictive validity and interrater reliability of the FACE-CARAS toolkit in a CAMHS setting. *Crim Behav Ment Health*. 2018;1–10. https://doi.org/10.1002/cbm.2104