BMJ Open

Single Best Answer multiple choice questions may give a false impression of competence: Cross-sectional study comparing Single Best Answer and Very Short Answer questions in 20 UK medical schools

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-032550
Article Type:	Research
Date Submitted by the Author:	26-Jun-2019
Complete List of Authors:	Sam, Amir; Imperial College Faculty of Medicine, Westacott, R; University of Leicester, Gurnell, Mark; University of Cambridge, Wellcome Trust-MRC Institute of Metabolic Science Wilson, Rebecca; Imperial College London Meeran, Karim; Imperial College Faculty of Medicine, Endocrinology Brown, Celia; The University of Warwick, Warwick Medical School (WMS)
Keywords:	MEDICAL EDUCATION & TRAINING, Assessment, applied medical knowledge

Single Best Answer multiple choice questions may give a false impression of competence: Cross-sectional study comparing Single Best Answer and Very Short Answer questions in 20 UK medical schools

Amir H. Sam[1], Rachel Westacott[2], Mark Gurnell[3], Rebecca K. Wilson[1], Karim Meeran[1] and Celia A. Brown (0000-0002-7526-0793)[4*]

1: School of Medicine, Imperial College London, South Kensington Campus, London, SW7 2AZ, UK.

Amir Sam: Head of Curriculum and Assessment Development

Karim Meeran: Director of Teaching

Rebecca Wilson: Clinical Education Fellow

2: Leicester Medical School, University of Leicester, Lancaster Road, Leicester, LE1 7HA, UK.

Rachel Westacott: Head of Assessment

3: University of Cambridge, Metabolic Research Laboratories, Welcome Trust-MRC Institute of Metabolic Science, Level 4 Box 289, Addenbrooke's Hospital, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, UK.

Mark Gurnell: Clinical SubDean and Professor of Clinical Endocrinology

4: Warwick Medical School, University of Warwick Coventry, CV4 7AL, UK.

Celia Brown: Associate Professor

* Corresponding author: email celia.brown@warwick.ac.uk, 00447813815762

Word count: 2,720

ABSTRACT

Objectives

To compare candidate performance between traditional best-of-five single-best-answer (SBA) questions and very short answer (VSA) questions, in which candidates must generate their own answers of between one and five words. The primary objective was to determine if the mean positive cue rate for SBAs exceeded the null hypothesis guessing rate of 20%.

Design

This was a cross-sectional study undertaken in 2018.

Setting

20 medical schools in the United Kingdom.

Participants

1,417 volunteer medical students preparing for their final undergraduate medicine examinations (total eligible population across all 20 participating medical schools approximately 4,700).

Interventions

Students completed a 50-question VSA test, followed immediately by the same test in SBA format, using a novel digital exam delivery platform which also facilitated rapid marking of VSAs.

Main outcome measures

The mean positive cue rate across SBAs: the percentage of students getting the SBA format of the question correct after getting the VSA format incorrect. Internal consistency, item discrimination and the pass rate using Cohen standard setting for VSAs and SBAs were also calculated and a cost analysis in terms of marking the VSA was performed.

Results

The study was completed by 1,417 students. Mean student scores were 21 percentage points higher for SBAs. The mean positive cue rate was 42.7% (95% CI 36.8 to 48.6%), one-sample t-test against <=20%: t=7.53, p<0.001. Internal consistency was higher for VSAs than SBAs and median item discrimination equivalent. The estimated marking cost was £2,655 ($3,500), with 24.5 hours of clinician time required (1.25 seconds per student per question).

Conclusions

SBA questions can give a false impression of students' competence. VSAs appear to have greater authenticity and can provide useful information regarding students' cognitive errors, helping to improve learning as well as assessment. Electronic delivery and marking of VSAs is feasible and cost-effective.

Strengths and limitations of the study

- This is the largest and only multi-centre study to date on the use of very-short-answer questions (VSAs) for the assessment of applied medical knowledge of medical students.
- A robust marking process for VSAs was used involving multiple markers and independent checking.
- Students volunteered to participate and the assessment was formative, so some responder bias is likely.
- Students did not spend long on the single-best-answer format as they had just read the questions in VSA format. This was to avoid cueing in the VSA but may have biased positive cue rates upwards.

INTRODUCTION

For many years single-best-answer (SBA) questions have been the cornerstone of written assessments testing applied medical knowledge,[1,2] including in high-stakes licensing assessments such as the US Medical Licensing Examination, the membership examinations of many UK Royal Colleges and graduation-level examinations of most UK medical schools. These questions consist of a clinical vignette, a lead-in question and (usually) five potential answers, one of which is the best answer (example in Box 1). Well-written SBAs can assess more than simple recall[3] and have a number of advantages: they are easy to mark electronically making scoring quick and accurate, they produce internally consistent measures of ability and they are acceptable to candidates because there is no inter-marker variability.[4,5] However, the provision of five possible answers means that a candidate may identify the correct answer by using cues provided in the option list or test-taking behaviours such as word-association.[2,6] Candidates may focus on practising exam technique rather than understanding the principles of the subject matter and honing their cognitive reasoning skills, thus adversely impacting learning behaviours.[6,7]

Because patients do not present with a list of five possible diagnoses, investigations or treatment options,[8] SBA questions do not simulate the "situations they [the candidates] will face when they undertake patient-related clinical tasks".[9,p. 66] Any alternative method of assessing applied medical knowledge must therefore provide increased content and response process validity, without resulting in significant reductions in other types of validity, reliability, acceptability, educational impact or an unacceptable increase in cost.[10] Very short answer (VSA) questions are a potential solution.[11,12] Like SBAs, VSAs have a clinical vignette followed by a lead-in question and can also be delivered electronically. Instead of having an answer list with the candidate being required to select one option, the candidate must provide their own answer. Questions are constructed so that the answer required is one to five words in length (example in Box 1). Pre-programmed correct and incorrect answers allow the VSA responses of most candidates to be marked automatically. Any

responses not fitting the pre-programmed answers are then reviewed by a team of clinicians who determine which should be accepted as correct. The software stores the additional correct and incorrect responses making each question much quicker and easier to mark if it is used in subsequent assessments.

Preliminary evidence suggests that VSAs have at least the same level of internal consistency as SBAs; they are practical, can be marked relatively quickly and may encourage positive changes in learning behaviours.[11,12] An electronic VSA exam platform has been developed by the UK Medical Schools Council Assessment Alliance to complement their existing SBA platform, which is already widely used by medical schools throughout the UK. We used this VSA platform to undertake a large, multi-centre cross-sectional study to evaluate VSAs in comparison to SBAs. In particular, our objective was to determine if validity is compromised by the provision of five answer options in SBAs by calculating 'positive cue' rates for each question. A 'positive cue' occurs when a student gives an incorrect answer in the VSA format but correctly answers the question in SBA format.[13] We also sought to determine if using VSAs had an impact on other aspects of assessment utility (reliability, potential educational impact and cost), as well as the ability of individual VSA and SBA questions to discriminate according to student performance on other questions.

Box 1: Example of a question in VSA and SBA format

A 60 year old man has 2 days of a swollen, painful right leg. He has a history of hypertension and takes ramipril. He is otherwise well.

He has a swollen right leg. The remainder of the examination is normal.

Investigations:

Haemoglobin	140 g/L	(130–175)
White cell count	8.0 × 10^9/L	(3.8–10.0)
Platelets	340 × 10^9/L	(150–400)
Creatinine	94 µmol/L	(60–120)
Total Calcium	2.5 mmol/L	(2.2–2.6)
Alanine Aminotransferase	30 IU/L	(10–50)
Alkaline Phosphatase	99 IU/L	(25–115)
APTT	30 seconds	(22–41)
Prothrombin Time	12 seconds	(10–12)

Urinalysis: normal

Chest X-ray: normal

Venous duplex ultrasound scan: thrombus in superficial femoral vein

What is the most appropriate additional investigation?

VSA answers marked as correct (total students answering correctly: N=33, 2.3%):

Variants of CT chest/abdomen/pelvis were accepted

Most common incorrect VSA Answers (N, % of all students):

CT Pulmonary Angiogram (487, 34%)

D-Dimer (386, 27%)

ECG (107, 7.6%)

Ankle Brachial Pressure Index (58, 4.1%)

SBA answer options (N, % of all students choosing each):

A. CT of abdomen and pelvis (957, 68%)

B. Serum carcinoembryonic antigen (57, 4.0%)

C. Serum prostate specific antigen (100, 7.1%)

D. Serum protein electrophoresis (143, 10%)

E. Ultrasonography of abdomen (157, 11%)

METHODS

Study population

All UK medical schools with graduation-level assessments (N=32) were invited to participate in this cross-sectional study. Assessment leads at schools agreeing to participate invited all of their final year students to participate and organised the delivery of the assessment within a 10 week window between September and November 2018. Participation in the study by both schools and students was voluntary; students were provided with information about the study prior to taking part. Completion of both assessments was taken as evidence of informed consent. The study was approved by the Imperial College of Medicine Medical Education Ethics Committee (reference MEEC1718-100).

Materials

We developed a 50-question formative assessment, using the same questions in both VSA and SBA formats (Appendix 1). Participants were first given two hours to complete the VSA

format and a further hour to complete the SBA format. Those entitled to extra time in summative assessments (e.g. those with dyslexia) were given an additional 30/15 minutes (25%). The assessments were completed under examination conditions in computer rooms at each medical school.

Marking and feedback

SBAs were marked electronically using a pre-determined answer key. Two clinicians (AS and RW) reviewed all of the answers for each VSA and coded each response as correct (scoring 1 mark) or incorrect (0 marks; any blank responses also scored 0). A third clinician (KM) was available to arbitrate any queries. A fourth clinician (RKW) subsequently reviewed all answers to check for any errors in marking. The time taken to mark each question was recorded.

Once all schools had completed the assessment, the SBA paper with answers and explanations was made available to all UK medical schools. Schools were informed of any questions in which <50% participating students answered the SBA question correctly for generic feedback but were not provided with individual student data. Students were able to review their individual performance in each assessment by logging into the exam platform.

Statistical analysis

The study administration team produced an Excel file containing answers and scores for each student for each question. Each student was allocated a numerical code and each school an alphabetical code before the data were sent to the research team to ensure anonymity. The data were transferred into Stata v15[14] for analysis.

For each participant/question combination, we identified whether providing answer options gave a positive cue. A positive cue occurred when a participant gave an incorrect answer to the VSA format of a question but the correct answer to the SBA.[13] We calculated the positive cue rate for each question as follows:

$$\text{Positive cue rate} = \frac{\text{Number of participants answering VSA incorrectly AND SBA correctly}}{\text{Number of participants answering VSA incorrectly}} \times 100$$

If all students answering the VSA incorrectly simply guessed at the SBA, the expected positive cue rate would be 20%. We therefore undertook a one-sided one-sample t-test against the null hypothesis that the rate would be <=20%, using a critical p-value of 0.025.

We also plotted the positive cue rate against VSA facility for each question to show how these statistics interact to enable identification of questions where poor knowledge (as assessed by the VSA) would be masked by the use of the equivalent SBA (questions with low VSA facility and a high positive cue rate).

Methods of analysis of additional outcomes are summarised in Table 1. Where statistical significance testing was undertaken in these additional analyses a critical p-value of <0.01 was used.

Sample size

A sample size calculation was undertaken in Stata v15. 47 questions would be required to detect a mean positive cue rate of >=30% (standard deviation 20%), in a one-sided one-sample t-test with alpha=0.02 and power=90%, against the null hypothesis value of <=20%.

Patient and public involvement (PPI)

There were no funds or time allocated for PPI so we were unable to involve patients.

RESULTS

The study was completed by 1,417 students from 20 UK medical schools; data from all participants was included in the analysis so there were no missing data. The range in student numbers between schools was 3 to 256, which was due to differences in cohort size as well as differences in participation rates. Data on participant characteristics and reasons for non-participation of schools and individual students were not collected. The mean time spent on each format of the assessment for students without extra time was 82/120 minutes

(SD 19 minutes) for the VSA and 24/60 minutes (SD 10 minutes) for the SBA, although students were reading the questions for the second time in SBA format. The mean score for the SBA items was 30.5/50 (SD 5.6) and that for the VSA items was 19.9 (SD 5.88).

Table 2 presents summary statistics comparing the SBA and VSA formats of the assessment (question-level data are shown in Appendix 2). The mean difference in question facility was 20 percentage points in favour of SBAs. The mean positive cue rate of 42.7% (95% CI 36.8 to 48.6%) was just over double the expected rate had all students answering the VSA format incorrectly taken a random guess at the SBA.

Figure 1 shows a scatter diagram of the positive cue rate against VSA facility. The diagram is split into four quadrants. The "concerning" top-left quadrant identifies questions where poor knowledge as assessed by the VSA (facility <0.5 or 50%) is masked by the use of the SBA: a high positive cue rate (>50%) leads to SBA facilities at least 25 percentage points above the VSA facility. There were 11 items in this quadrant (22%), as summarised in Table 3.

Questions in the top-right quadrant of Figure 1 (N=7/50, 14%) have a high positive cue rate (>50%), but the SBA format does not conceal a major cohort-level deficit in knowledge because the VSA facility was also fairly high (>0.5). Those in the bottom-left quadrant (N=24, 48%) have a low VSA facility (<0.5), but a lack of knowledge amongst the cohort is also revealed with the SBA format as the positive cue rate is low (<50%). Finally, questions in the bottom-right quadrant (N=8, 16%) have high VSA facility (>0.5) and a low positive cue rate (<50%).

The internal consistency of the VSA format of the assessment (Cronbach's alpha 0.731) was higher than for the SBA format (0.693); this difference was statistically significant: $F_{1416,1416}=1.262$, $p<0.001$. Median question discrimination was 0.184 for SBAs and 0.192 for VSAs; this difference was not statistically significant ($z=-1.36$, $p=0.175$).

In terms of potential educational impact, the Kappa statistic of 0.59 ($p<0.01$) suggests "moderate" agreement between pass/fail decisions on the two assessments using the Cohen method of standard setting, albeit with a much lower pass mark for the VSA paper. Despite a strong positive correlation between participants' scores on the two formats ($r=0.822$, $p<0.001$), 161 students (11.4%) would have passed the SBA but failed the VSA whereas 92 students (6.5%) would have passed the VSA but failed the SBA.

The two primary question markers worked together, each spending a total of 8 hours and 34 minutes marking the 50 VSAs. The median time per question per marker was 9:43 minutes, with an inter-quartile range of 5:00 to 13:09 and overall range 1:55 to 25:39), and the distribution was highly positively skewed. The third clinician on-hand to arbitrate spent a total of 30 minutes doing so. To mitigate marker bias, all marking was subsequently checked by a fourth marker, who spent a total of 6 hours and 57 minutes doing so. Assuming all markers were at consultant level, the total marking time cost, for this 50 question paper for 1,417 students was £2,655 ($3,500).

DISCUSSION

Our findings highlight the advantages of using VSAs rather than SBAs to assess applied clinical knowledge in high stakes summative medical exams. VSA scores are a better representation of students' unprompted level of knowledge, with the average student scoring 21 percentage points lower on the VSA version of the assessment. If the questions used in our study are representative of undergraduate medical curricula and average question difficulty, then cues provided in SBAs could impact on the validity of at least one quarter of the examination. These items are assessing the candidate's ability to use cues or engage in test-taking behaviours such using the answer options to make deductions about the correct answer rather than using clinical reasoning, arriving at the correct answer by eliminating wrong SBA answer options[8] and/or 'best-guessing' from the answers available. We have shown that VSAs mitigate this risk by removing the option menu and compelling candidates

to determine the correct answer themselves based on the clinical information provided which is more akin to clinical practice. Linked to this, an added benefit of the VSA format is its ability to help identify deficits in students' knowledge and/or cognitive reasoning. The themes of the questions with high positive cue rates and low VSA facility highlight areas of the curriculum where students lack understanding, and where using the SBA format can therefore provide a false measure of students' competence. Importantly, several VSA questions highlighted significant cognitive errors, which were not apparent in their SBA counterparts, or indeed even considered as possible student responses by the person authoring the question. The question in Box 1 is a good example: a venous thromboembolism has been confirmed therefore rendering a D-dimer irrelevant, yet just over a quarter of students chose this option in the VSA. More concerning, just over one-third of students would have ordered a CT pulmonary angiogram in a patient with no respiratory symptoms or signs, thereby exposing the patient to a significant dose of unnecessary radiation without any likely therapeutic benefit. It is also possible that further investigation to exclude an occult malignancy would not have been instituted.

VSAs were non-inferior to SBAs on other indices of assessment utility. In terms of feasibility, the electronic delivery platform functioned well and participating medical schools did not report any problems associated with delivering the assessment. The platform also facilitated remote marking. VSAs are more time-consuming to mark than SBAs, but not prohibitively so. The marking time for an individual VSA will fall significantly with repeated use as pre-existing marking schemes are re-applied. VSAs also had slightly higher internal consistency (a measure of reliability) and comparable question discrimination, as seen in previous small scale pilot studies.[11,12]

This study involved 20 medical schools across the United Kingdom. The large number of medical schools that took part in the study and the overall high number of participants makes this the largest study comparing VSAs with SBAs and suggests that the findings of this study are generalizable across the UK and potentially internationally. Previous studies have

highlighted the benefits and shortcomings of SBA questions,[2,4-7,13], but our work provides large-scale empirical data to test some of these claims using an alternative and feasible question format as the comparator.

Our study has several limitations. Students volunteered to participate and therefore there is likely to be some responder bias. This assessment was formative and was sat at variable timeframes ahead of students' medical school summative assessments (depending on the individual dates for summative assessments which varied for each participating medical school). Students are therefore likely to have prepared and participated in a different way than for a summative assessment, especially as for some schools, final exams were several months after the study date. Students all sat the SBA questions after the VSA questions to ensure there was no cueing in the VSA. This means that positive cue rates may have been biased upwards because participants had a second look at the questions during the SBA paper, which may have contributed to them arriving at the correct answer along with having the answer options.

We have not yet undertaken a criterion-based approach to standard setting using expert judgment, so are unable to determine whether the full cueing effect of SBAs is accounted for in common standard setting processes such as Angoff[15] or Ebel.[16] Furthermore, this study was also not designed to evaluate all components of assessment utility including acceptability to stakeholders. The previous smaller-scale pilots of VSAs reported that students found VSAs more challenging, but appreciated the additional validity they offered.[11,12]

Key extensions to this work should include the study of how SBA and VSA questions are standard set relative to performance and a comparison of the predictive validity of SBA and VSA scores, particularly using measures of performance in clinical settings.

CONCLUSION

VSAs appear to provide a more accurate measure of a candidate's knowledge than SBAs. They also offer greater insight into cognitive errors, thereby offering opportunities to hone teaching, feedback and learning, as well as creating summative assessments with greater validity. Unlike short-answer questions, modified essay formats or clinical reasoning problems,[9] VSAs are straightforward to deliver in an electronic format and efficient to mark. We need to know that medical students and trainees have the required applied medical knowledge to practice safely without test scores being confounded by the ability to use the cues of SBA answer options. Our results suggest that VSAs could provide a more authentic method of assessing medical knowledge whilst maintaining most of the cost-efficiency of SBAs.

Figure legend

Figure 1: Scatter diagram of VSA facility and the positive cue rate

Legend to quadrants in Figure 1:

Top-left: N=11 questions with low VSA facility (<0.5 or 50%) and a high positive cue rate (>50%).

Top-right: N=7 questions with high VSA facility (>0.5) and a high positive cue rate (>50%).

Bottom-left: N=24 questions with low VSA facility (<0.5) and a low positive cue rate (<50%).

Bottom-right: N=8 questions with high VSA facility (>0.5) and a low positive cue rate (<50%).

Acknowledgements

We would like to thank the Medical Schools Council Assessment Alliance (MSCAA) management and administration team who supported this study. The MSCAA Board provided support and helpful comments on the study protocol. This study required the support of Assessment leads at participating schools, invigilators and students, who all volunteered their time.

Data sharing

The individual item-level data for each student participant are not available.

Transparency declaration

AS (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that there were no discrepancies from the study as originally planned.

Authors' contributions

The study was designed and implemented by AS, RWe, CT and MG. AS and RWe wrote the question paper. AS, RWe and KM undertook the initial marking, which was verified by RWi. CT undertook the data analysis and wrote the first draft of the paper, supported by RWe for the literature review. AS, RWe, MG, KM and RWi provided critical comments on the paper during each round of drafting.

Funding and role of the funder

The MSCAA supported this study through the provision of staff time and travel costs, but no financial payments were made. The MSCAA Board provided suggestions on study design but had no role in data analysis, interpretation and did not input into the writing of this paper, although they did see the draft prior to submission. As noted in the acknowledgements section, MSCAA staff assisted with data collection.

MG is supported by the NIHR Cambridge Biomedical Research Centre.

CB is supported by the NIHR CLAHRC West Midlands initiative. This paper presents independent research and the views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Conflicts of interest

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare the following, in addition to the funding section above:

CB, AS, MG and RW are elected members of the MSCAA Board.

MG and AS are Advisors for the GMC UK Medical Licensing Assessment.

Dissemination

The results of this study have been reported to the participating medical schools. Participating medical students have received feedback on their performance in the assessment. They will have access to the study results on publication of this article.

REFERENCES

1. Coderre SP, Harasym P, Mandin H, Fick G. The impact of two multiple-choice question formats on the problem-solving strategies used by novices and experts. *BMC medical education.* 2004;4(1):23.
2. Heist BS, Gonzalo JD, Durning S, Torre D, Elnicki DM. Exploring clinical reasoning strategies and test-taking behaviors during clinical vignette style multiple-choice examinations: a mixed methods study. *Journal of graduate medical education.* 2014;6(4):709-714.
3. Bibler Zaidi NL, Grob KL, Monrad SM, Kurtz JB, Tai A, Ahmed AZ, et al. Pushing Critical Thinking Skills With Multiple-Choice Questions: Does Bloom's Taxonomy Work? *Academic Medicine.* 2018;93(6):856-859.
4. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet.* 2001;357(9260):945-949.
5. Pugh D, De Champlain A, Touchie C. Plus ÇA Change, plus C'Est Pareil: Making a Continued Case for the Use of MCQs in Medical Education. *Medical teacher.* 2018:1-9.
6. Raduta C. Consequences the extensive use of multiple-choice questions might have on student's reasoning structure. *Rom. Journ. Phys.* 2013;58:1363-1380.
7. McCoubrie P. Improving the fairness of multiple-choice questions: a literature review. *Medical teacher.* 2004;26(8):709-712.
8. Surry LT, Torre D, Durning SJ. Exploring examinee behaviours as validity evidence for multiple-choice question examinations. *Medical education.* 2017;51(10):1075-1085.
9. ten Cate O, Durning SJ. Approaches to Assessing the Clinical Reasoning of Preclinical Students. In: ten Cate O, Custers E, Durning S, eds. *Principles and Practice of Case-based Clinical Reasoning Education*: Springer, Cham; 2018:65-72.

10. Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ.* 1996;1(1):41-67.
11. Sam AH, Field SM, Collares CF, van der Vleuten CP, Wass VJ, Melville C, et al. Very-short-answer questions: reliability, discrimination and acceptability. *Medical education.* 2018;52(4):447-455.
12. Sam AH, Hameed S, Harris J, Meeran K. Validity of very short answer versus single best answer questions for undergraduate assessment. *BMC medical education.* 2016;16(1):266.
13. Schuwirth L, Van der Vleuten C, Donkers H. A closer look at cueing effects in multiple-choice questions. *Medical education.* 1996;30(1):44-49.
14. StataCorp. *Stata Statistical Software: Release 15.* College Station, Texas: LP S; 2016.
15. Angoff W. Scales, norms and equivalent scores. In: Thorndike R, ed. *Educational Measurement*. Washington, DC: American Council on Education; 1971:508-600.
16. Ebel RL. Procedures for the analysis of classroom tests. *Educ Psychol Meas.* 1954;14:352-364.
17. Feldt L. A test of the hypothesis that Cronbach's alpha reliability coefficient is the same for two tests administered to the same sample. *Psychometrika.* 1980;45(1):99-105.
18. Holsgrove G. *Reliability issues in the assessment of small cohorts.* London: General Medical Council; 2010.
19. Curtis L, Burns A. *Unit costs of health and social care 2018.* Canterbury: Kent Uo; 2018.
20. Cohen-Schotanus J, van der Vleuten C. A standard setting method with the best performing students as point of reference: Practical and affordable. *Medical teacher.* 2010;32(2):154-160.

Table 1: Additional data analysis

Component of assessment utility being evaluated	**Method of analysis**
Reliability: internal consistency	Cronbach's alpha coefficient for each type of question compared using the method of Feldt;[17] the Spearman-Brown formula was then used to estimate the number of questions of each type required for an alpha of 0.8.[18]
Cost: time taken to mark VSAs	The total minutes of consultant time required to mark the VSA, costed at the 2016/17 hourly rate for a hospital consultant (including on-costs and overheads) of £108[19] ($143).
Potential educational Impact: effect on pass/fail rates	Cohen standard setting[20] applied to both VSA and SBA total scores; pass/fail decisions for the two assessments were then compared using Cohen's Kappa.
Question discrimination	Pearson correlation coefficient (point-biserial) between students' scores on each question and those on all other questions combined (item-rest correlation) for each type of question; the difference between question types was compared using a Wilcoxon signed rank sum test (for paired, skewed data).

Table 2: Comparison of SBA and VSA questions and scores

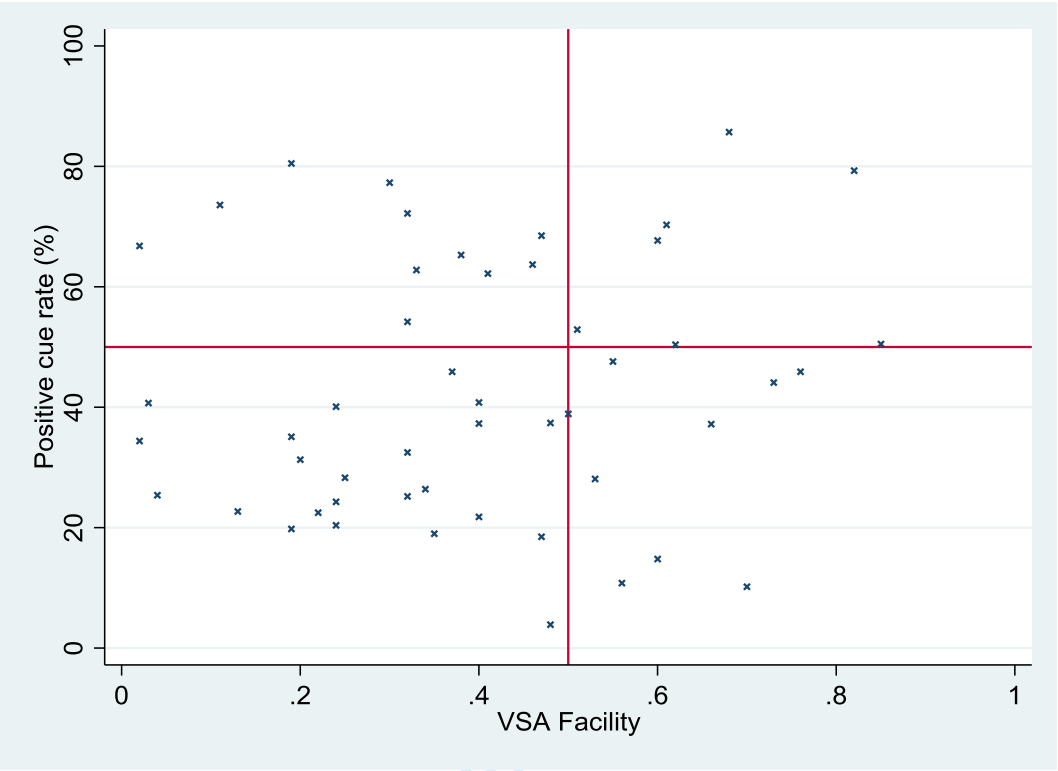
	SBA	**VSA**	**SBA – VSA difference and Statistical significance**
Question facility[a] Mean (SD), Range	0.61 (0.20), 0.16 to 0.95	0.40 (0.21), 0.02 to 0.85	0.21 (0.19), -0.32 to 0.65 Paired t-test t=7.89, p<0.001
Positive cue rate (question level) Mean (SD), Range (%)	42.7 (21.3), 3.9 to 85.7		One-sample t-test (NH=<20%) t=7.53, p<0.001
Internal consistency (Cronbach's alpha)	0.693	0.731	-0.038 $F_{1416,1416}$=1.262, p<0.001
Questions required for an alpha of 0.8	89	74	15
Cohen pass mark[b]	28/50	18/50	N/A
Pass rate using Cohen pass mark (%)	71.2	66.3	Kappa = 0.59 z=22.2, p<0.001
Question discrimination Median (IQR), Range	0.184 (0.135 to 0.220), 0.003 to 0.287	0.192 (0.121 to 0.259). -0.006 to 0.395	-0.004 (-0.083 to 0.034), -0.296 to 0.225 Wilcoxon test z=-1.36, p=0.175

[a] Facility: proportion of students answering correctly.

[b] Calculated as 60% of the score of the 95th percentile student and assuming scores due to guessing of 20% for the SBA and 0% for the VSA.

Table 3: Question statistics and themes of questions with VSA facility <0.5 and positive cue rate>50%

Question	SBA Facility	VSA Facility	Difference	Positive cue rate (%)	Theme
9	0.84	0.19	0.65	80.5	Investigations of diabetes insipidus
41	0.82	0.30	0.52	77.3	Diagnosis of cerebellar stroke
3	0.76	0.11	0.65	73.6	Assessment of patient following house fire
25	0.80	0.32	0.48	72.2	Treatment of delirium
16	0.80	0.47	0.34	68.5	Investigations of a neck lump
4	0.68	0.02	0.65	66.8	Further investigation of unprovoked DVT
43	0.78	0.38	0.40	65.3	Determining Glasgow Coma Scale (GCS) Score
13	0.79	0.46	0.32	63.7	Diagnosis of headache
21	0.74	0.33	0.41	62.8	Causative organism of malaria
8	0.76	0.41	0.35	62.2	Diagnosis of (o)esophageal rupture
31	0.66	0.32	0.34	54.2	Management of gout



Appendix 1: Question paper

1. A 73 year old man collapses on the surgical ward 24 hours after having a sigmoid volvulus reduced by sigmoidoscopy.

He has no pulse and an ECG shows asystole. Chest compressions and ventilation are started. The cardiac arrest team are with the patient.

Which is the most appropriate next step in management?

A. Cardiac defibrillation
B. Intravenous 0.9% sodium chloride
C. Intravenous adrenaline/epinephrine
D. Intravenous atropine sulfate
E. Transcutaneous pacing

Correct Answer(s): C

Justification for correct answer

Epinephrine the only recommended treatment for asystole. Atropine no longer recommended.

2. A 78 year old woman has pain in both shoulders, hips and thighs. She is very stiff on waking in the morning and takes 2–3 hours to loosen up. She finds getting dressed difficult.

Investigations:
Erythrocyte sedimentation rate 67 mm/hr (<20)
CRP 87 mg/L (<5)

What is the most appropriate initial treatment?

A. Co-codamol
B. Leflunomide
C. Methotrexate
D. Naproxen
E. Prednisolone

Correct Answer(s): E

Justification for correct answer

The scenario describes typical features of polymyalgia rheumatica, which responds well to prednisolone and poorly to analgesics such as Naproxen or cocodamol. No features to suggest inflammatory arthritis

3. A 36 year old man is rescued from a house fire.

He is alert and talking but has a dull headache. His pulse rate is 98 bpm, BP 139/86 mmHg, respiratory rate 22 breaths per minute and oxygen saturation 100% breathing 15 L/min oxygen via a non-rebreather mask.

Which is the most appropriate parameter to measure?

A. Bicarbonate
B. Carboxyhaemoglobin
C. Haemoglobin
D. Lactate
E. Methaemoglobin

Correct Answer(s): B

Justification for correct answer

Carbon monoxide inhalation likely. Will not show on pulse oximeter as is read as oxyhaemoglobin

4. A 60 year old man has 2 days of a swollen, painful right leg. He has a history of hypertension and takes ramipril. He is otherwise well.

He has a swollen right leg. The remainder of the examination is normal.

Investigations:

Haemoglobin	140 g/L	(130–175)
White cell count	8.0×10^9/L	(3.8–10.0)
Platelets	340×10^9/L	(150–400)
Creatinine	94 µmol/L	(60–120)
Calcium	2.5 mmol/L	(2.2–2.6)
ALT	30 IU/L	(10–50)
ALP	99 IU/L	(25–115)
APTT	30 seconds	(22–41)
PT	12 seconds	(10–12)

Urinalysis: normal

Chest X-ray: normal

Venous duplex ultrasound scan: thrombus in superficial femoral vein

Which is the most appropriate additional investigation?

A. CT of abdomen and pelvis
B. Serum carcinoembryonic antigen
C. Serum prostate specific antigen
D. Serum protein electrophoresis
E. Ultrasonography of abdomen

Correct Answer(s): A

Justification for correct answer

The patient has an unprovoked DVT. Patients should be offered CT scan abdomen and pelvis to help identify possible malignancy.

5. A 65 year old woman had a mechanical aortic valve replacement and coronary revascularisation 3 days ago. She is being treated with dalteparin sodium. She is also taking aspirin long term.

Which is the most appropriate long-term patient management?

A. Apixaban
B. Clopidogrel
C. Continue dalteparin sodium
D. Rivaroxaban
E. Warfarin sodium

Correct Answer(s): E

Justification for correct answer

All patients with mechanical valves require treatment with aspirin and warfarin. Low molecular weight heparin is used as bridging anti-coagulation but not long-term. There is no place for novel oral anticoagulants (yet).

6. The association between maternal smoking during pregnancy and low birthweight can be studied by obtaining smoking histories from pregnant women at the time of first prenatal visit, then assessing birthweight at delivery and analysing the data according to the smoking histories.

Which is the best description of this type of study?

A. Case–control
B. Cross-sectional
C. Prospective cohort
D. Randomised clinical trial
E. Retrospective cohort

Correct Answer(s): C

Justification for correct answer

This is a prospective cohort study where a group of individuals who differ with respect to one or more factors are followed, to determine how these factors affect outcomes. A case-control study finds individuals with a given outcome along with a suitable control group and looks back retrospectively at how many individuals from both groups had the exposure(s) of interest. A cross-sectional study measures prevalence in a population at a given point in time. A randomised clinical trial tests the effects of a specific intervention or interventions against a control group. A retrospective cohort begins after the outcome has occurred and looks back at the exposure of interest.

7. An 18 year old woman has had a chronic skin condition for several years. She has noticed that she gets well-defined areas of scale formation at the sites of minor skin injuries, such as scratches or insect bites, typically when the injuries are healing. The scaling can persist for several weeks or months.

What is the most likely underlying skin condition?

A. Acne vulgaris
B. Eczema
C. Psoriasis
D. Seborrhoeic dermatitis
E. Vitiligo

Correct Answer(s): C

Justification for correct answer

This is typical of Koebnerisation. Psoriasis is by far the commonest underlying cause, though it can also occur in vitiligo.

8. A 64 year old man has vomiting and severe chest pain after eating a large meal.

His temperature is 37.6°C, pulse rate 130 bpm, BP 95/50 mmHg and respiratory rate 30 breaths per minute. There is palpable subcutaneous emphysema on the left side of his neck.

Which is the most likely diagnosis?

A. Diaphragmatic rupture
B. Mallory–Weiss tear
C. Necrotising fasciitis
D. Oesophageal rupture
E. Spontaneous pneumohaemothorax

Correct Answer(s): D

Justification for correct answer

The scenario describes Boerhaave's syndrome (oesophageal rupture).

9. A 54 year old woman has polyuria and the feeling that she needs to drink continuously. She drinks at least 1 litre of water before bedtime and gets up three to four times during the night to pass urine. She has another glass of water each time that she gets up.

Investigations:
Sodium 140 mmol/L (135–146)
Potassium 4.1 mmol/L (3.5–5.3)
Urea 4.5 mmol/L (2.5–7.8)
Creatinine 86 µmol/L (60–120)
Calcium 2.56 mmol/L (2.2–2.6)
Fasting glucose 4.8 mmol/L (3.0–6.0)
Serum osmolality 295 mOsmol/kg (285–295)
Urinary osmolality 86 mOsmol/kg (100–1000)

After 8 hours of a water deprivation test, the serum osmolality is 308 mOsmol/kg and the urinary osmolality is 152 mOsmol/kg.

Following the administration of desmopressin, the serum osmolality is 286 mOsmol/kg and the urinary osmolality is 660 mOsmol/kg.

Which is the most appropriate next investigation?

A. CT scan of thorax, abdomen and pelvis
B. MR scan of pituitary
C. Oral glucose tolerance test
D. Technetium-99 Sestamibi parathyroid scan
E. Supervised fluid restriction and daily weights

Correct Answer(s): B

Justification for correct answer

The test results are consistent with cranial diabetes insipidus.

10. A 30 year old woman has irregular periods, decreased libido and galactorrhoea.

Visual field examination is normal.

Investigations:
Prolactin 5000 mU/L (100–500)
Pregnancy test: negative

MR scan of pituitary shows a 4-mm mass in the sella turcica.

Which is the most appropriate management?

A. Cabergoline
B. Dexamethasone
C. Octreotide
D. Radiotherapy
E. Transsphenoidal surgery

Correct Answer(s): A

Justification for correct answer

Cabergoline (a dopamine agonist) is the first-line treatment for a micro- and macroprolatinomas.#

11. A 23 year old man is admitted to the acute surgical ward with appendicitis and is prepared for theatre. Although he has not eaten for 24 hours, he has been vomiting on and off all day.

Which airway device provides protection for the lungs from regurgitated stomach contents?

A. Guedel (oral) airway
B. i-gel ® (supraglottic) airway
C. Laryngeal mask airway
D. Nasopharyngeal airway
E. Tracheal tube

Correct Answer(s): E

Justification for correct answer

i-gels are often used in cardiac arrest situations as they are easier to place than tracheal tubes. However, only the tracheal tube can seal the trachea off and protect against aspiration.

12. A 32 year old man is referred to a gastroenterology clinic with hepatomegaly. He has a history of type 2 diabetes that is diet controlled. He is taking no regular medication. He is a non-smoker and drinks approximately 16 units of alcohol per week. He is married with no children.

Cardiovascular and respiratory examinations are normal. His abdomen is soft, with a 3 cm palpable liver edge. His BMI is 23 kg/m^2(18–25).

Investigations:

Albumin	38 g/L	(35–50)
ALT	90 IU/L	(10–50)
ALP	112 IU/L	(25–115)
Bilirubin	15 µmol/L	(<17)
Ferritin	710 µg/L	(12–200)
CRP	6 mg/L	(<5)

Which is the most appropriate next investigation?

A. Hepatitis C serology
B. Liver biopsy
C. Reticulocyte count
D. Serum γGT
E. Transferrin saturation

Correct Answer(s): E

Justification for correct answer

Transferrin saturation is the screening test for haemochromatosis.

13. A 30 year old woman has severe headache 24 hours after a spinal anaesthetic.

Her temperature is 37.1°C, pulse rate 90 bpm and BP 120/80 mmHg.

Which is the most likely diagnosis?

A. Low pressure headache
B. Meningitis
C. Migraine
D. Subarachnoid haemorrhage
E. Subdural haemorrhage

Correct Answer(s): A

Justification for correct answer

Most likely cause in this setting - common phenomenon.

14. A 27 year old man is brought to the Emergency Department with left-sided chest pain of sudden onset that is worse on taking a deep breath.

His temperature is 36.8°C, pulse rate 126 bpm, BP 108/60 mmHg, respiratory rate 28 breaths per minute and oxygen saturation 94% breathing air.

Investigations:
ECG: sinus tachycardia

Which is the most appropriate next investigation?

A. Chest X-ray
B. CT pulmonary angiography
C. D dimers
D. Echocardiography
E. Ventilation/perfusion isotope lung scan

Correct Answer(s): A

Justification for correct answer

CXR to rule out other pathology before Well's score and then CTPA (or V/Q) if PE likely or D-dimer if PE unlikely. (NICE CG144) Diagnostic investigations for pulmonary embolism 1.1.7 If a patient presents with signs or symptoms of pulmonary embolism (PE), carry out an assessment of their general medical history, a physical examination and a chest X ray to exclude other causes. [2012]

15. A healthy 23 year old man is scheduled to undergo an elective arthroscopy of his knee. He is to have a general anaesthetic for the operation and asks the pre-operative assessment nurse how long he needs to fast beforehand.

Which are the most appropriate fasting times for clear liquids and solids respectively?

A. 2 h for clear liquids, 2 h for solids
B. 2 h for clear liquids, 6 h for solids
C. 2 h for clear liquids, 12 h for solids
D. 6 h for both clear liquids and solids
E. 6 h for clear liquids, 12 h for solids

Correct Answer(s): B

Justification for correct answer

http://www.aagbi.org/sites/default/files/Perioperative_fasting_in_adults_and_children_.4.pdf Standard national guidelines for elective patients with no problems affecting gastric emptying. Too long a period of fasting is unnecessary whilst residual solid food in the stomach poses a big risk of aspiration/asphyxiation.

16. A 28 year old woman presents to her GP with a neck lump that she noticed incidentally when rubbing her neck.

There is a smooth, non-tender 1.5 cm mobile lump within the thyroid gland.

Investigations:
TSH 2.3 mU/L (0.3–4.2)
Free T4 17 pmol/L (9–25)

Which is the most appropriate next investigation?

A. CT of neck
B. No further investigations
C. Thyroid antibodies
D. Thyroid scintigraphy
E. Ultrasonography of neck

Correct Answer(s): E

Justification for correct answer

This is a non-functional thyroid nodule so needs ultrasound to classify - FNA may then be indicated.

17. A 78 year old man has type 2 diabetes. His clinician does not invite him to join an internet-based self-monitoring programme because she considers him to be too old to engage with it effectively.

What is the most appropriate description of the clinician's approach?

A. Bias
B. Discrimination
C. Inequity
D. Prejudice
E. Stereotyping

Correct Answer(s): B

Justification for correct answer

Discrimination is the unjust or prejudicial treatment of different categories of people.

18. A 47 year old man with hypertension attends for annual review. He takes ramipril (10 mg once daily).

His BP is 138/78 mmHg.

Investigations:

Sodium	139 mmol/L	(135–146)
Potassium	6.2 mmol/L	(3.5–5.3)
Urea	5.0 mmol/L	(2.5–7.8)
Creatinine	90 µmol/L	(60–120)

Which is the most appropriate immediate action?

A. Add indapamide
B. Advise low potassium diet
C. Change ramipril to amlodipine
D. Reduce dose of ramipril
E. Repeat urea and electrolytes

Correct Answer(s): E

Justification for correct answer

This is likely spurious - and needs repeat.

19. A 33 year old woman has 4 months of joint pain and stiffness, predominantly affecting her feet. This is worst in the morning and gradually improves through the day. She feels tired all the time but reports no other health problems.

Which investigation would confirm the most likely diagnosis?

A. Anti-cyclic citrillinated peptide antibody
B. Anti-double-stranded DNA antibodies
C. Antinuclear antibody
D. CRP
E. Rheumatoid factor

Correct Answer(s): A

Justification for correct answer

The correct answer is A (Anti CCP) antibody. This is the test that has the highest specificity for rheumatoid arthritis, which is what the question is asking. CCP antibodies are found in 80% of people with rheumatoid arthritis, but fewer than 0.5% of healthy individuals. Rheumatoid factor is present in up to 10% of the healthy population, and whilst of similar sensitivity to CCP, is much less specific. CRP is a non-specific marker of inflammation, and can often be normal in early rheumatoid. ANA testing is high sensitivity (but low specificity) test for connective tissue disorders such as SLE and Sjorgren's.

20. A 27 year old man has severe central chest pain. He admits to using cocaine shortly before the onset of the chest pain, but says that he had used it on only two previous occasions.

He is distressed and sweating. His pulse rate is 115 bpm and BP 118/68 mmHg. An ECG shows sinus tachycardia with ST elevation in the lateral leads, and several ventricular ectopics.

Which is the mechanism by which cocaine has caused this acute episode?

A. Blockade of myocyte repolarisation
B. Coronary artery spasm
C. Enhanced platelet aggregation
D. Increased systemic vascular resistance
E. Rupture of pre-existing arterial plaques

Correct Answer(s): B

Justification for correct answer

Cocaine-induced ACS.

21. An 18 year old man, who was born in the UK, develops drowsiness and confusion 2 days after returning from visiting his grandparents in Malawi. Over the past week he has had recurrent episodes of high fever.

Investigations:Haemoglobin 92 g/L (130–175)
White cell count 3.2×10^9/L (3.8–10.0)
Platelets 184×10^9/L (150–400)
Blood film parasites visible

Which is the most likely causative organism?

A. Plasmodium falciparum
B. Plasmodium malariae
C. Plasmodium vivax
D. Trypanosoma brucei
E. Trypanosoma cruzi

Correct Answer(s): A

Justification for correct answer

Malaria is the most likely diagnosis. Cerebral involvement makes falciparum more likely. T cruzi is seen in S America. T brucei could cause these symptoms but is much rarer.

22. A 31 year old man visits his GP with a painless lump in his scrotum.

There is a well-defined, non-tender spherical 1 cm mass on the right side of the scrotum. It is superior to the testis and transilluminates.

Which is the most likely diagnosis?

A. Abscess
B. Epididymal cyst
C. Hydrocoele
D. Inguinal hernia
E. Testicular tumour

Correct Answer(s): B

Justification for correct answer

Separate and can get above mass which is cystic is an epididymal cyst (spermatocele) which is benign and usually asymptomatic and managed conservatively. Although a hydrocele trans illuminates it would surround the testis.

23. A 68 year old man has 3 days of worsening vomiting and abdominal pain. He has not passed any stool for 3 days. He has a history of a right hemicolectomy for Dukes' A (T1, N0) bowel cancer 6 months ago.

He is dehydrated and his abdomen is distended.

Which is the most likely diagnosis?

A. Adhesional small bowel obstruction
B. Anastomotic leak
C. Cholecystitis
D. Pancreatitis
E. Tumour recurrence

Correct Answer(s): A

Justification for correct answer

Adhesional bowel obstruction is most likely as he has had a hemicolectomy. Recurrence is less likely because this is Dukes A (early stage). Too late after surgery for anastamotic leak. Pancreatitis and cholecystitis are less likely because they don't cause constipation.

24. A 78 year old patient is admitted with chronic oropharyngeal dysphagia. He has left ventricular systolic dysfunction from ischaemic heart disease. He is breathless on exertion, particularly when climbing stairs. He is being prepared for a percutaneous endoscopic gastrostomy feeding tube. The passage of a nasogastric tube has been unsuccessful, and he is having nil by mouth.

He weighs 70 kg. His pulse rate is 72 bpm and BP 132/80 mmHg.

Which is the correct volume of maintenance fluids (in mL) to prescribe for the next 24 hours?

A. 1750
B. 2450
C. 2800
D. 3000
E. 3250

Correct Answer(s): A

Justification for correct answer

Maintenance fluid requirements for someone with underlying cardiac disease is recommended as 20-25 mL/kg. This gives an upper volume of 25 x 70= 1750. This conforms to current NICE guideline CG 174 (2013).

25. A 75 year old woman is reviewed 4 days after a fractured neck of femur repair. She has been agitated and upset, particularly at night. She has punched nurses and keeps trying to leave the ward. She has seen strange men in black capes entering the ward and believes that they are controlling the hospital. When she was seen in the memory clinic 6 months ago, she was found to have mild cognitive impairment.

What is the most appropriate treatment?

A. Chlordiazepoxide hydrochloride
B. Chlorpromazine hydrochloride
C. Donepezil hydrochloride
D. Haloperidol
E. Memantine hydrochloride

Correct Answer(s): D

Justification for correct answer

This woman has delirium. Therefore, low dose haloperidol would be the best option, according to NICE guidelines.

26. A 72 year old woman has 6 months of increasing pain at the base of her right thumb. She is having difficulty opening jars and sewing. She is otherwise well. No other joints are painful. She is taking regular analgesia.

The first carpometacarpal joint is swollen and tender, with reduced opposition of the thumb.

Which is the most likely diagnosis?

A. De Quervain's tenosynovitis
B. Gout
C. Osteoarthritis
D. Rheumatoid arthritis
E. Septic arthritis

Correct Answer(s): C

Justification for correct answer

This is a classic description of osteoarthritis and a common site Gout and septic arthritis would have a much more acute history. Rheumatoid arthritis would affect multiple joints. The pain of de Quervains tenosynovitis would be felt over the radial aspect of the wrist. Rheumatoid arthritis usually affects multiple joints.

27. A 62 year old man has acute breathlessness with a weak cough, following a recent viral upper respiratory infection. Over the past 4 months, he has had double vision, limb weakness and slurred speech when tired.

His respiratory rate is 18 breaths per minute and oxygen saturation 96% breathing air. He is sweating and using his accessory muscles of inspiration.

Which is the most appropriate test to monitor his respiratory function?

A. Arterial blood gas
B. FEV_1
C. FVC
D. Peak expiratory flow rate
E. Ratio of FEV_1 to FVC

Correct Answer(s): C

Justification for correct answer

Myasthenic crisis is an acute respiratory failure characterised by forced vital capacity (FVC) below 1 L, negative inspiratory force (NIF) of 20 cm H2O or less, and the need for ventilatory support. The use of accessory muscles indicates significant inspiratory weakness. Weak cough indicates weakness of expiratory muscles. Arterial blood gas analysis commonly shows hypercapnia before hypoxia. There should be a low threshold for endotracheal intubation due to rapid deterioration of bulbar and respiratory muscles.

28. A 40 year old man has 4 days of left flank pain associated with fever, nausea and vomiting.

His temperature is 39.6°C, pulse rate 118 bpm and BP 90/40 mmHg. Imaging shows an obstructing proximal left ureteric stone with severe hydronephrosis.

He is treated with intravenous antibiotics and intravenous fluids.

Which is the most appropriate next step in management?

A. Lithotripsy
B. Nephrostomy
C. Retrograde pyelography
D. Ureteric stent
E. Urethral catheter

Correct Answer(s): B

Justification for correct answer

The renal pelvis should be decompressed with a nephrostomy.

29. A 19 year old man has 2 days of right-sided chest pain and breathlessness on exertion. He smokes cannabis and takes cocaine.

His temperature is 36.4°C, pulse rate 108 bpm, BP 112/80 mmHg, respiratory rate 24 breaths per minute and oxygen saturation 94% breathing air. His trachea is central. He has reduced breath sounds at the right apex.

What is the most likely diagnosis?

A. Acute coronary syndrome
B. Coronary artery spasm
C. Pneumonia
D. Pneumothorax
E. Pulmonary embolism

Correct Answer(s): D

Justification for correct answer

Pneumothoraces are more commonly encountered in young, tall men who smoke. There is no past medical history provided in the vignette confirming the presence of existing lung disease - this excludes secondary pneumothorax as a diagnosis.

30. An 80 year old woman is admitted to the Emergency Department after being found collapsed at home. She has central chest pain.

Her pulse rate is 30 bpm, BP 70/40 mmHg and respiratory rate 26 breaths per minute. Her 12-lead ECG shows sinus bradycardia with no evidence of myocardial ischaemia.

Which is the most appropriate initial treatment?

A. Adrenaline/epinephrine
B. Atropine sulfate
C. Dobutamine
D. Normal saline
E. Permanent cardiac pacemaker

Correct Answer(s): B

Justification for correct answer

Atropine is the first line initial treatment for sinus bradycardia.

31. A 62 year old man develops acute pain, redness, swelling and warmth of his right first metatarsophalangeal joint. He has a history of gout and hypertension. His medications are allopurinol, amlodipine and ramipril.

His eGFR is >60 mL/min/1.73 m^2(>60).

Which is the most appropriate next step in his management?

A. Change allopurinol to febuxostat
B. Start naproxen
C. Stop allopurinol
D. Stop amlodipine
E. Stop ramipril

Correct Answer(s): B

Justification for correct answer

The patient has acute gout. The immediate management would be to commence an NSAID.

32. A previously healthy 10 year old boy has deafness of new onset. He has a history of a recent respiratory tract infection.

Tuning fork tests show:
a) when the tuning fork is placed in the middle of his forehead he hears the tone loudest in his right ear;
b) when the tuning fork is held in front of his right external auditory meatus it is quieter than when it is placed on the bone behind the same ear;
c) when the tuning fork is held in front of the left external auditory meatus the sound is louder than when it is placed on the bone behind the same ear.

Which ear(s) is/are affected and which type of hearing loss is this?

A. Bilateral mixed deafness
B. Left conductive deafness
C. Left sensorineural deafness
D. Right conductive deafness
E. Right sensorineural deafness

Correct Answer(s): D

Justification for correct answer

The combination of lateralisation of Weber's test to the right and a negative Rinne's test on the right (i.e. bone conduction louder than air conduction) occurs with right conductive deafness. This also fits with the clinical history of a recent respiratory tract infection.

33. A 55 year old man has 2 days of painful red swelling of his left lower leg. He has a history of type 2 diabetes mellitus and takes metformin.

His temperature is 37.6°C. He has a tender erythematous area extending from the ankle to the proximal calf.

What is the most likely causative organism?

A. Bacteroides species
B. Proteus mirabilis
C. Pseudomonas aeruginosa
D. Staphylococcus epidermidis
E. Streptococcus pyogenes

Correct Answer(s): E

Justification for correct answer

Streptococcus is the most common pathogen in leg cellulitis (including in patients with diabetes).

34. A 55 year old woman has a tender, erythematous, swollen hard cord in the long saphenous vein distribution in her calf. She has a longstanding history of bilateral varicose veins.

An ultrasound scan shows superficial thrombophlebitis without deep vein thrombosis.

Which is the most appropriate treatment?

A. Dipyridamole
B. Flucloxacillin
C. Naproxen
D. Paracetamol
E. Rivaroxaban

Correct Answer(s): C

Justification for correct answer

NSAIDs are the first-line treatment for superficial thrombophlebitis (NICE CKS thrombophlebitis – superficial, May 2017)

35. A 64 year old woman is due to undergo an arthroscopy of her knee. She has type 2 diabetes and takes metformin (500 mg twice daily) and gliclazide (80 mg each morning).

Her glycated haemoglobin is 54 mmol/mol (20–42). She is scheduled first on the morning day case list and is asked to fast from midnight the previous night.

Which is the most appropriate plan for managing her diabetic medication?

A. Continue both drugs and start a variable-rate insulin infusion
B. No change to usual medication
C. Omit both drugs and start a variable-rate insulin infusion
D. Omit gliclazide and continue metformin
E. Omit metformin and continue gliclazide

Correct Answer(s): D

Justification for correct answer

Joint British Diabetes Societies' 2016 guidelines - Management of adults with diabetes undergoing surgery and elective procedures: Improving standards. Principles are to minimise fasting times (hence first on list) and disruption to usual medication. In this setting where the procedure is fairly minor, patient should be eating again at lunchtime and hence there is no call for insulin in any form. Specifics are in Appendix 2 of document. https://www.diabetes.org.uk/resources-s3/2017-09/Surgical%20guidelines%202015%20-%20full%20FINAL%20amended%20Mar%202016_0.pdf

36. A 67 year old man has difficulty walking. He states that he has to raise his left leg higher in the air than normal to avoid scraping his toes on the ground when he walks.

When he raises the left foot from the floor, the ankle assumes a plantar-flexed position with the toes directed towards the floor.

Which nerve is most likely to be affected?

A. Common peroneal
B. Medial plantar
C. Saphenous
D. Superficial peroneal
E. Tibial

Correct Answer(s): A

Justification for correct answer

The patient has foot drop due to loss of active dorsiflexion. The muscles affected are supplied by the common peroneal nerve.

37. A 35 year old woman has 6 months of cyclical pain in both breasts. She has recently noticed a lump in the right breast. There is diffuse nodularity of the axillary tails of both breasts with a discrete 20 mm mass in the upper outer quadrant of the right breast. Fine needle aspiration is performed and 5 mL of brown fluid is removed, with disappearance of the mass. Cytology shows cellular debris with no malignant cells.

What is the most likely diagnosis?

A. Breast abscess
B. Breast carcinoma
C. Fat necrosis
D. Fibroadenoma
E. Fibrocystic disease

Correct Answer(s): E

Justification for correct answer

Fibrocystic disease characteristically causes pain associated with the menstrual cycle. The fine needle aspiration supports this with no malignant cells seen. A fibroadenoma is a solid lump. Pain if present would be localised with fat necrosis. Breast abscess would be a more acute history and again would not be expected to cause bilateral breast pain.

38. A 70 year old man has a brief episode of twitching that starts in his left hand and spreads up the arm over 2 minutes, then stops. His arm feels weak for an hour afterwards. He had an ischaemic stroke affecting his left side 6 months ago with good functional recovery. He has a history of type 2 diabetes mellitus and is taking clopidogrel, metformin, ramipril and simvastatin. He is anxious about a further stroke.

There is no weakness on neurological examination.

Which is the most likely diagnosis?

A. Functional episode
B. Hypoglycaemia
C. Migraine
D. Partial seizure
E. Right hemisphere transient ischaemic attack

Correct Answer(s): D

Justification for correct answer

The description fits with partial seizure affecting his right hemisphere as a result of a previous stroke.

39. A 63 year old woman has 4 months of abdominal bloating, fatigue and nausea.

She is found to have with ovarian cancer. Staging CT is performed to look for lymphatic spread and metastatic disease.

To what regional lymph nodes is her tumour most likely to spread initially?

A. Deep inguinal nodes
B. External iliac nodes
C. Internal iliac nodes
D. Para-aortic nodes
E. Superficial inguinal nodes

Correct Answer(s): D

Justification for correct answer

The main lymphatic drainage of the ovary is to the para-aortic nodes. The iliac nodes are less frequently involved.

40. A 52 year old woman has had three episodes of severe epigastric pain associated with vomiting over the past 3 months. The episodes occurred following eating and lasted for about 1 hour. She has type 2 diabetes mellitus and takes metformin.

Abdominal examination is normal. Her BMI is 35 kg/m^2(18–25).

Investigations:

ALT	15 IU/L	(10–50)
ALP	71 IU/L	(25–115)
Bilirubin	9 µmol/L	(<17)

Ultrasound scan of abdomen: single 2-cm gallstone in gallbladder, common bile duct normal, evidence of fatty liver.

Which is the most appropriate management?

A. Endoscopic retrograde cholangiopancreatography
B. Laparoscopic cholecystectomy
C. MR cholangiopancreatography
D. Open cholecystectomy
E. Ursodeoxycholic acid

Correct Answer(s): B

Justification for correct answer

The patient has symptomatic gallstone disease and laparoscopic cholecystectomy is indicated. Percutaneous cholecystostomy may be used in patients who are not fit for surgery. urosdeoxycholic acid may be used for gallstone dissolution, but is not part of common UK practice.

41. A 56 year old woman develops vertigo, nausea, vomiting and intense occipital headache of sudden onset. She is unable to walk without falling. She has a history of hypertension treated with ramipril.

Her temperature is 37.4°C, pulse rate 94 bpm, BP 146/92 mmHg, respiratory rate 12 breaths per minute and oxygen saturation 96% breathing air. She has multidirectional nystagmus and some clumsiness of her right arm.

Which is the most likely diagnosis?

A. Acute labyrinthitis
B. Benign paroxysmal positional vertigo
C. Cerebellar stroke
D. Ménière's disease
E. Multiple sclerosis

Correct Answer(s): C

Justification for correct answer

Triad of headache, nausea/vomitting and ataxia. Profound imbalance, sudden onset and prominent headache suggest cerebellar stroke.

42. A 79 year old woman has 3 months of a left leg venous ulcer that is slowly healing.

Ankle–brachial pressure Indices are 0.9 on the left side and 1.1 on the right side (0.8–1.2).

Which is the most appropriate management?

A. Compression stockings
B. Diagnostic biopsy
C. Full-length graduated compression bandaging
D. No further management required
E. Repeat ankle–brachial pressure indices in 3 months

Correct Answer(s): A

Justification for correct answer

Despite gradual improvement this venous ulcer would be best managed with stockings, which would also help prevent further lesions.

43. A 26 year old woman sustains a head injury in a motorcycle accident. Her eyes are closed, but she opens them when asked to do. She is confused about what happened and about where she is, but attempts to talk about it. She is repeatedly attempting to remove the cannula from her right wrist.

What is her GCS?

A. 3
B. 7
C. 9
D. 12
E. 14

Correct Answer(s): D

Justification for correct answer

Opens eyes in response to voice = 3 Confused, disoriented = 4 Localizes painful stimuli = 5.

44. A 65 year old man has sudden pain and redness in his right eye. He also has a headache and nausea.

Visual acuity is 6/60 in the right eye. The eye is congested, with a hazy cornea and mid-dilated pupil.

Which is the most likely diagnosis?

A. Acute glaucoma
B. Conjunctivitis
C. Corneal ulcer
D. Scleritis
E. Uveitis

Correct Answer(s): A

Justification for correct answer

All the symptoms and signs described can occur with acute glaucoma. Uveitis whilst causing red eye, headache and visual disturbance is associated with a small pupil. Scleritis, corneal ulcer and conjunctivitis are not generally associated with headache and nausea or a significant drop in visual acuity.

45. A 45 year old woman attends her GP surgery with symptoms of vaginal soreness, itching and discharge. She has had recurrent episodes of vaginal candidiasis over the past 4 months.

The labia minora are red and swollen. A diagnosis of vaginal candidiasis is made.

What is the most appropriate investigation at this stage?

A. Glycated haemoglobin
B. HIV test
C. Sexually transmitted infection screen
D. Test her partner for candidiasis
E. Vaginal pH testing

Correct Answer(s): A

Justification for correct answer

Recurrent candidiasis indicates the need to test for diabetes mellitus.

46. A 68 year old man collapses when rising from a chair and is seen in the emergency department 45 minutes later. He is conscious but has reduced power in his left arm and leg (3/5 and 4/5 respectively) and is slurring his speech. He has a past medical history of COPD and hypertension. He smokes 10 cigarettes per day.

He has bilateral scattered wheeze and carotid bruits on auscultation.

Which is the most appropriate initial radiological investigation?

A. Carotid ultrasonography
B. Cerebral angiography
C. CT cerebral venography
D. CT of head
E. MR imaging of brain

Correct Answer(s): D

Justification for correct answer

This is a probable CVA (presentation and risk factors). Due to rapid presentation he is a candidate for systemic thrombolysis. Non-contrast CT head is the most rapid investigation to exclude intracranial haemorrhage and allow thrombolysis.

47. A 25 year old man sustains 40% full-thickness burns in a house fire. Despite intensive treatment, he becomes breathless and hypotensive. He develops a petechial rash.

His temperature is 38°C, pulse rate 110 bpm, BP 80/50 mmHg and oxygen saturation 96% breathing 40% oxygen.

Investigations:

Haemoglobin	110 g/L	(130–175)
White cell count	4.2×10^9/L	(3.8–10.0)
Platelets	15×10^9/L	(150–400)
APTT	75 seconds	(22–41)
PT	25 seconds	(10–12)
Fibrinogen	0.7 g/dL	(1.5–4.0)
Fibrinogen degradation products	137 mg/mL	(<8)

Which is the most likely diagnosis?

A. Anaphylactic reaction to antibiotics
B. Disseminated intravascular coagulation
C. Fat embolism
D. Immune thrombocytopenic purpura
E. Pulmonary embolism

Correct Answer(s): B

Justification for correct answer

The history, signs and investigations are all most indicative of DIC.

48. A 65 year old woman with newly diagnosed advanced lung cancer has 1 day of breathlessness and 1 week of progressive headache.Her pulse rate is 88 bpm, respiratory rate 20 breaths per minute and oxygen saturation 95% breathing 4 L/min oxygen via nasal prongs. She has a swollen face and neck and distended veins on her chest. Her chest is clear. CT scan of chest shows mediastinal lymphadenopathy compressing the superior vena cava.

Which is the most appropriate initial treatment?

A. Intravenous alteplase
B. Intravenous dexamethasone
C. Intravenous heparin infusion
D. Intravenous mannitol
E. Insert endovenous stent

Correct Answer(s): B

Justification for correct answer

The patient has superior vena cava obstruction. The standard initial treatment is dexamethasone to reduce tumour swelling. There is no evidence of thrombus to justify systemic anticoagulation and systemic thrombolysis has no place. Intravenous mannitol is not used. It is worth noting that although dexamethasone is commonly recommended there is limited controlled data confirming its effectiveness. Insertion of endovenous stent would be considered if there was stridor, but would likely follow intubation and steroids. The following guideline is of use: http://www.palliativecareguidelines.scot.nhs.uk/guidelines/palliative-emergencies/Superior-Vena-Cava-Obstruction.aspx

49. An 80 year old man presents to the Emergency Department with dizziness and melaena of recent onset. He has a metallic mitral valve and is taking aspirin and warfarin sodium. His pulse rate is 80 bpm and BP 122/70 mmHg.

Investigations:

Haemoglobin	105 g/L	(130–175)
White cell count	7.0×10^9/L	(3.8–10.0)
Platelets	676×10^9/L	(150–400)
INR	9.6	(1.0)

He is treated with intravenous vitamin K and is blood cross-matched.

What is the most appropriate additional treatment?

A. Fresh frozen plasma
B. Pantoprazole
C. Protamine sulfate
D. Prothrombin complex concentrate
E. Tranexamic acid

Correct Answer(s): D

Justification for correct answer

Prothrombin complex concentrate is used to reverse warfarin in medical emergencies. It is quicker to administer than FFP and can reverse anti-coagulation within minutes. FFP also carries the risk of allergic reactions, transfusion-related lung injury and volume overload. PCC is therefore considered first-line to reverse warfarin. The other drugs do not reverse warfarin.

50. A 65 year old man has 3 weeks of progressive ankle oedema. He is a lifelong heavy smoker and drinks 12 units of alcohol per week.

His BP is 125/85 mmHg and oxygen saturation 98% breathing air. He has marked bilateral pitting ankle oedema.

Investigations:

Creatinine	85 µmol/L	(60–120)
Urinary protein: creatinine ratio	400 mg/mmol	(<30)
Fasting glucose	5.7 mmol/L	(3.0–6.0)
Total cholesterol	9 mmol/L	(<5.0) Albumin 20 g/L (35–50)

He is treated with furosemide.

Which investigation is most likely to be diagnostic?

A. Chest X-ray
B. Renal arteriography
C. Renal auto-antibody screen
D. Renal biopsy
E. Serum protein electrophoresis

Correct Answer(s): D

Justification for correct answer

Nephrotic syndrome in adults requires renal biopsy to identify the cause, prior to definitive treatment.

Appendix 2: Question-level Statistics

N=1,417 students	Facility		Discrimination (Item-rest correlation)		
Question	SBA	VSA	SBA	VSA	Positive cue rate (%)
1	0.62	0.40	0.22	0.23	40.8
2	0.66	0.50	0.27	0.19	38.9
3	0.76	0.11	0.14	0.26	73.6
4	0.68	0.02	0.03	0.12	66.8
5	0.76	0.66	0.23	0.19	37.2
6	0.76	0.62	0.14	0.18	50.4
7	0.82	0.76	0.16	0.12	45.9
8	0.76	0.41	0.25	0.40	62.2
9	0.84	0.19	0.19	0.20	80.5
10	0.39	0.35	0.13	0.21	19.0
11	0.62	0.53	0.13	0.23	28.1
12	0.42	0.03	0.20	0.11	40.7
13	0.79	0.46	0.21	0.34	63.7
14	0.50	0.56	0.07	0.00	10.8
15	0.53	0.32	0.20	0.19	32.5
16	0.80	0.47	0.16	0.13	68.5
17	0.42	0.19	0.12	0.14	35.1
18	0.35	0.02	0.15	0.00	34.4
19	0.60	0.40	0.18	0.25	37.3
20	0.84	0.60	0.22	0.23	67.7
21	0.74	0.33	0.18	0.23	62.8
22	0.34	0.19	0.19	0.22	19.8
23	0.94	0.68	0.01	0.12	85.7
24	0.40	0.25	0.15	0.11	28.3
25	0.80	0.32	0.13	0.14	72.2

N=1,417 students	Facility		Discrimination (Item-rest correlation)		
Question	SBA	VSA	SBA	VSA	Positive cue rate (%)
26	0.56	0.60	0.15	0.17	14.8
27	0.16	0.48	0.26	0.04	3.9
28	0.38	0.22	0.28	0.26	22.5
29	0.82	0.73	0.18	0.18	44.1
30	0.48	0.32	0.29	0.27	25.2
31	0.66	0.32	0.21	0.14	54.2
32	0.71	0.70	0.26	0.26	10.2
33	0.28	0.04	0.00	-0.01	25.4
34	0.31	0.13	0.11	0.09	22.7
35	0.37	0.24	0.08	0.05	20.4
36	0.76	0.51	0.18	0.30	52.9
37	0.64	0.55	0.20	0.14	47.6
38	0.49	0.34	0.23	0.29	26.4
39	0.49	0.40	0.19	0.31	21.8
40	0.61	0.48	0.18	0.17	37.4
41	0.82	0.30	0.24	0.05	77.3
42	0.30	0.47	0.05	0.35	18.5
43	0.78	0.38	0.13	0.12	65.3
44	0.87	0.61	0.19	0.34	70.3
45	0.41	0.24	0.23	0.16	24.3
46	0.89	0.85	0.20	0.20	50.5
47	0.95	0.82	0.12	0.29	79.3
48	0.53	0.24	0.17	0.25	40.1
49	0.43	0.20	0.20	0.34	31.3
50	0.65	0.37	0.25	0.32	45.9

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2-3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	7-8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	7-8
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	7
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	8-9
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	7-9
Bias	9	Describe any efforts to address potential sources of bias	12-13
Study size	10	Explain how the study size was arrived at	9
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	8-9
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8-9 Table 1
		(*b*) Describe any methods used to examine subgroups and interactions	N/A
		(*c*) Explain how missing data were addressed	9
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	N/A
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	9
		(b) Give reasons for non-participation at each stage	9
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	9
		(b) Indicate number of participants with missing data for each variable of interest	9
Outcome data	15*	Report numbers of outcome events or summary measures	9-11

			Table 2
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	10
		(*b*) Report category boundaries when continuous variables were categorized	N/A
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	10-11
Discussion			
Key results	18	Summarise key results with reference to study objectives	11-12
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	12-13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	14
Generalisability	21	Discuss the generalisability (external validity) of the study results	12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	16

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Cross-sectional study comparing Single Best Answer and Very Short Answer questions for the assessment of applied medical knowledge in 20 UK medical schools

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-032550.R1
Article Type:	Research
Date Submitted by the Author:	01-Aug-2019
Complete List of Authors:	Sam, Amir; Imperial College Faculty of Medicine, Westacott, R; University of Leicester, Gurnell, Mark; University of Cambridge, Wellcome Trust-MRC Institute of Metabolic Science Wilson, Rebecca; Imperial College London Meeran, Karim; Imperial College Faculty of Medicine, Endocrinology Brown, Celia; The University of Warwick, Warwick Medical School (WMS)
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Research methods
Keywords:	MEDICAL EDUCATION & TRAINING, Assessment, applied medical knowledge

Cross-sectional study comparing Single Best Answer and Very Short Answer questions for the assessment of applied medical knowledge in 20 UK medical schools

Amir H. Sam[1], Rachel Westacott[2], Mark Gurnell[3], Rebecca K. Wilson[1], Karim Meeran[1] and Celia A. Brown (0000-0002-7526-0793)[4]*

1: School of Medicine, Imperial College London, South Kensington Campus, London, SW7 2AZ, UK.

Amir Sam: Head of Curriculum and Assessment Development

Karim Meeran: Director of Teaching

Rebecca Wilson: Clinical Education Fellow

2: Leicester Medical School, University of Leicester, Lancaster Road, Leicester, LE1 7HA, UK.

Rachel Westacott: Head of Assessment

3: University of Cambridge, Metabolic Research Laboratories, Welcome Trust-MRC Institute of Metabolic Science, Level 4 Box 289, Addenbrooke's Hospital, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, UK.

Mark Gurnell: Clinical SubDean and Professor of Clinical Endocrinology

4: Warwick Medical School, University of Warwick Coventry, CV4 7AL, UK.

Celia Brown: Associate Professor

* Corresponding author: email celia.brown@warwick.ac.uk, 00447813815762

Word count: 3,362

ABSTRACT

Objectives

To compare candidate performance between traditional best-of-five single-best-answer (SBA) questions and very short answer (VSA) questions, in which candidates must generate their own answers of between one and five words. The primary objective was to determine if the mean positive cue rate for SBAs exceeded the null hypothesis guessing rate of 20%.

Design

This was a cross-sectional study undertaken in 2018.

Setting

20 medical schools in the United Kingdom.

Participants

1,417 volunteer medical students preparing for their final undergraduate medicine examinations (total eligible population across all UK medical schools approximately 7,500).

Interventions

Students completed a 50-question VSA test, followed immediately by the same test in SBA format, using a novel digital exam delivery platform which also facilitated rapid marking of VSAs.

Main outcome measures

The mean positive cue rate across SBAs: the percentage of students getting the SBA format of the question correct after getting the VSA format incorrect. Internal consistency, item discrimination and the pass rate using Cohen standard setting for VSAs and SBAs were also calculated and a cost analysis in terms of marking the VSA was performed.

Results

The study was completed by 1,417 students. Mean student scores were 21 percentage points higher for SBAs. The mean positive cue rate was 42.7% (95% CI 36.8 to 48.6%), one-sample t-test against <=20%: t=7.53, p<0.001. Internal consistency was higher for VSAs than SBAs and median item discrimination equivalent. The estimated marking cost was £2,655 ($3,500), with 24.5 hours of clinician time required (1.25 seconds per student per question).

Conclusions

SBA questions can give a false impression of students' competence. VSAs appear to have greater authenticity and can provide useful information regarding students' cognitive errors, helping to improve learning as well as assessment. Electronic delivery and marking of VSAs is feasible and cost-effective.

Strengths and limitations of the study

- This is the largest and only multi-centre study to date on the use of very-short-answer questions (VSAs) for the assessment of applied medical knowledge of medical students.
- A robust marking process for VSAs was used involving multiple markers and independent checking.
- Students volunteered to participate and the assessment was formative, so some responder bias is likely.
- Students did not spend long on the single-best-answer format as they had just read the questions in VSA format. This was to avoid cueing in the VSA but may have biased positive cue rates upwards.

INTRODUCTION

For many years single-best-answer (SBA) questions have been the cornerstone of written assessments testing applied medical knowledge,[1,2] including in high-stakes licensing assessments such as the US Medical Licensing Examination, the membership examinations of many UK Royal Colleges and graduation-level examinations of most UK medical schools. These questions consist of a clinical vignette, a lead-in question and (usually) five potential answers, one of which is the best answer (example in Box 1). Well-written SBAs can assess more than simple recall[3] and have a number of advantages: they are easy to mark electronically making scoring quick and accurate, they produce internally consistent measures of ability and they are acceptable to candidates because there is no inter-marker variability.[4,5] However, the provision of five possible answers means that a candidate may identify the correct answer by using cues provided in the option list or test-taking behaviours such as word-association.[2,6] Candidates may focus on practising exam technique rather than understanding the principles of the subject matter and honing their cognitive reasoning skills, thus adversely impacting learning behaviours.[6,7]

Because patients do not present with a list of five possible diagnoses, investigations or treatment options,[8] SBA questions do not simulate the "situations they [the candidates] will face when they undertake patient-related clinical tasks".[9,p. 66] Any alternative method of assessing applied medical knowledge must therefore provide increased content and response process validity, without resulting in significant reductions in other types of validity, reliability, acceptability, educational impact or an unacceptable increase in cost.[10] Very short answer (VSA) questions are a potential solution.[11,12] Like SBAs, VSAs have a clinical vignette followed by a lead-in question and can also be delivered electronically. Instead of having an answer list with the candidate being required to select one option, the candidate must provide their own answer. Questions are constructed so that the answer required is one to five words in length (example in Box 1). Pre-programmed correct and incorrect answers allow the VSA responses of most candidates to be marked automatically. Any

responses not fitting the pre-programmed answers are then reviewed by a team of clinicians who determine which should be accepted as correct. The software stores the additional correct and incorrect responses making each question much quicker and easier to mark if it is used in subsequent assessments.

Preliminary evidence suggests that VSAs have at least the same level of internal consistency as SBAs; they are practical, can be marked relatively quickly and may encourage positive changes in learning behaviours.[11,12] An electronic VSA exam platform has been developed by the UK Medical Schools Council Assessment Alliance to complement their existing SBA platform, which is already widely used by medical schools throughout the UK. We used this VSA platform to undertake a large, multi-centre cross-sectional study to evaluate VSAs in comparison to SBAs. In particular, our objective was to determine if validity is compromised by the provision of five answer options in SBAs by calculating 'positive cue' rates for each question. A 'positive cue' occurs when a student gives an incorrect answer in the VSA format but correctly answers the question in SBA format.[13] We also sought to determine if using VSAs had an impact on other aspects of assessment utility (reliability, potential educational impact and cost), as well as the ability of individual VSA and SBA questions to discriminate according to student performance on other questions.

Box 1: Example of a question in VSA and SBA format

A 60 year old man has 2 days of a swollen, painful right leg. He has a history of hypertension and takes ramipril. He is otherwise well.

He has a swollen right leg. The remainder of the examination is normal.

Investigations:

Haemoglobin	140 g/L (130–175)
White cell count	8.0 × 10^9/L (3.8–10.0)
Platelets	340 × 10^9/L (150–400)
Creatinine	94 µmol/L (60–120)
Total Calcium	2.5 mmol/L (2.2–2.6)
Alanine Aminotransferase	30 IU/L (10–50)
Alkaline Phosphatase	99 IU/L (25–115)
APTT	30 seconds (22–41)
Prothrombin Time	12 seconds (10–12)

Urinalysis: normal

Chest X-ray: normal

Venous duplex ultrasound scan: thrombus in superficial femoral vein

What is the most appropriate additional investigation?

VSA answers marked as correct (total students answering correctly: N=33, 2.3%):

Variants of CT chest/abdomen/pelvis were accepted

Most common incorrect VSA Answers (N, % of all students):

CT Pulmonary Angiogram (487, 34%)

D-Dimer (386, 27%)

ECG (107, 7.6%)

Ankle Brachial Pressure Index (58, 4.1%)

SBA answer options (N, % of all students choosing each):

A. CT of abdomen and pelvis (957, 68%)

B. Serum carcinoembryonic antigen (57, 4.0%)

C. Serum prostate specific antigen (100, 7.1%)

D. Serum protein electrophoresis (143, 10%)

E. Ultrasonography of abdomen (157, 11%)

METHODS

Study population

All UK medical schools with graduation-level assessments (N=32) were invited to participate in this cross-sectional study. Assessment leads at schools agreeing to participate invited all of their final year students to participate and organised the delivery of the assessment within a 10 week window between September and November 2018. Participation in the study by both schools and students was voluntary; students were provided with information about the study prior to taking part. Completion of both assessments was taken as evidence of informed consent. The study was approved by the Imperial College of Medicine Medical Education Ethics Committee (reference MEEC1718-100).

Materials

We developed a 50-question formative assessment, using the same questions in both VSA and SBA formats (Appendix 1). Participants were first given two hours to complete the VSA

format and a further hour to complete the SBA format. Those entitled to extra time in summative assessments (e.g. those with dyslexia) were given an additional 30/15 minutes (25%). The assessments were completed under examination conditions in computer rooms at each medical school.

Marking and feedback

SBAs were marked electronically using a pre-determined answer key. VSA marking is semi-automated; the electronic platform checks the student's response against a list of pre-determined answers. Those responses that match this list are automatically marked as correct. Two clinicians (AS and RW) reviewed all the remaining answers for each VSA and coded each response as correct (scoring 1 mark) or incorrect (0 marks; any blank responses also scored 0). A third clinician (KM) was available to arbitrate any queries. A fourth clinician (RKW) subsequently reviewed all answers to check for any errors in marking. The time taken to mark each question was recorded.

Once all schools had completed the assessment, the SBA paper with answers and explanations was made available to all UK medical schools. Schools were informed of any questions in which <50% participating students answered the SBA question correctly for generic feedback but were not provided with individual student data. Students were able to review their individual performance in each assessment by logging into the exam platform.

Statistical analysis

The study administration team produced an Excel file containing answers and scores for each student for each question. Each student was allocated a numerical code and each school an alphabetical code before the data were sent to the research team to ensure anonymity. The data were transferred into Stata v15[14] for analysis.

For each participant/question combination, we identified whether providing answer options gave a positive cue. A positive cue occurred when a participant gave an incorrect answer to

the VSA format of a question but the correct answer to the SBA.[13] We calculated the positive cue rate (as a percentage) for each question as follows:

$$\text{Positive cue rate} = \frac{\text{Number of participants answering VSA incorrectly AND SBA correctly}}{\text{Number of participants answering VSA incorrectly}} \times 100$$

If all students answering the VSA incorrectly simply guessed at the SBA, the expected positive cue rate would be 20%. We therefore undertook a one-sided one-sample t-test against the null hypothesis that the rate would be <=20%, using a critical p-value of 0.025.

We also plotted the positive cue rate against VSA facility for each question to show how these statistics interact to enable identification of questions where poor knowledge (as assessed by the VSA) would be masked by the use of the equivalent SBA (questions with low VSA facility and a high positive cue rate).

Methods of analysis of additional outcomes are summarised in Table 1. Where statistical significance testing was undertaken in these additional analyses a critical p-value of <0.01 was used.

Sample size

A sample size calculation was undertaken in Stata v15. 47 questions would be required to detect a mean positive cue rate of >=30% (standard deviation 20%), in a one-sided one-sample t-test with alpha=0.02 and power=90%, against the null hypothesis value of <=20%.

Patient and public involvement (PPI)

There were no funds or time allocated for PPI so we were unable to involve patients.

RESULTS

The study was completed by 1,417 students from 20 UK medical schools (approximately 20% of all final year students); data from all participants was included in the analysis so there were no missing data (and we assumed any blank responses had been left intentionally blank and were scored 0). The range in student numbers between schools was

3 to 256 (median 45, inter-quartile range 21 to 103), which was due to differences in cohort size as well as differences in participation rates. Data on participant characteristics and reasons for non-participation of schools and individual students were not collected. The mean time spent on each format of the assessment for students without extra time was 82/120 minutes (SD 19 minutes) for the VSA and 24/60 minutes (SD 10 minutes) for the SBA, although students were reading the questions for the second time in SBA format. The mean score for the SBA items was 30.5/50 (SD 5.6) and that for the VSA items was 19.9 (SD 5.88).

Table 2 presents summary statistics comparing the SBA and VSA formats of the assessment (question-level data are shown in Appendix 2). The mean difference in question facility was 20 percentage points in favour of SBAs. The mean positive cue rate of 42.7% (95% CI 36.8 to 48.6%) was just over double the expected rate had all students answering the VSA format incorrectly taken a random guess at the SBA.

Figure 1 shows a scatter diagram of the positive cue rate against VSA facility. The diagram is split into four quadrants. The “concerning” top-left quadrant identifies questions where poor knowledge as assessed by the VSA (facility <0.5 or 50%) is masked by the use of the SBA: a high positive cue rate (>50%) leads to SBA facilities at least 25 percentage points above the VSA facility. There were 11 items in this quadrant (22%), as summarised in Table 3.

Questions in the top-right quadrant of Figure 1 (N=7/50, 14%) have a high positive cue rate (>50%), but the SBA format does not conceal a major cohort-level deficit in knowledge because the VSA facility was also fairly high (>0.5). Those in the bottom-left quadrant (N=24, 48%) have a low VSA facility (<0.5), but a lack of knowledge amongst the cohort is also revealed with the SBA format as the positive cue rate is low (<50%). Finally, questions in the bottom-right quadrant (N=8, 16%) have high VSA facility (>0.5) and a low positive cue rate (<50%).

The internal consistency of the VSA format of the assessment (Cronbach's alpha 0.731) was higher than for the SBA format (0.693); this difference was statistically significant: $F_{1416,1416}=1.262$, $p<0.001$. Median question discrimination was 0.184 for SBAs and 0.192 for VSAs; this difference was not statistically significant ($z=-1.36$, $p=0.175$).

In terms of potential educational impact, the Kappa statistic of 0.59 ($p<0.01$) suggests "moderate" agreement between pass/fail decisions on the two assessments using the Cohen method of standard setting, albeit with a much lower pass mark for the VSA paper. Despite a strong positive correlation between participants' scores on the two formats ($r=0.822$, $p<0.001$), 161 students (11.4%) would have passed the SBA but failed the VSA whereas 92 students (6.5%) would have passed the VSA but failed the SBA.

The two primary question markers worked together, each spending a total of 8 hours and 34 minutes marking the 50 VSAs. The median time per question per marker was 9:43 minutes, with an inter-quartile range of 5:00 to 13:09 and overall range 1:55 to 25:39), and the distribution was highly positively skewed. The third clinician on-hand to arbitrate spent a total of 30 minutes doing so. To mitigate marker bias, all marking was subsequently checked by a fourth marker, who spent a total of 6 hours and 57 minutes doing so. Assuming all markers were at consultant level, the total marking time cost, for this 50 question paper for 1,417 students was £2,655 ($3,500).

DISCUSSION

Our findings highlight the advantages of using VSAs rather than SBAs to assess applied clinical knowledge in high stakes summative medical exams. VSA scores are a better representation of students' unprompted level of knowledge, with the average student scoring 21 percentage points lower on the VSA version of the assessment. If the questions used in our study are representative of undergraduate medical curricula and average question difficulty, then cues provided in SBAs could impact on the validity of at least one quarter of the examination. These items are assessing the candidate's ability to use cues or engage in

test-taking behaviours such using the answer options to make deductions about the correct answer rather than using clinical reasoning, arriving at the correct answer by eliminating wrong SBA answer options[8] and/or 'best-guessing' from the answers available. We have shown that VSAs mitigate this risk by removing the option menu and compelling candidates to determine the correct answer themselves based on the clinical information provided which is more akin to clinical practice. Linked to this, an added benefit of the VSA format is its ability to help identify deficits in students' knowledge and/or cognitive reasoning. The themes of the questions with high positive cue rates and low VSA facility highlight areas of the curriculum where students lack understanding, and where using the SBA format can therefore provide a false measure of students' competence. Importantly, several VSA questions highlighted significant cognitive errors, which were not apparent in their SBA counterparts, or indeed even considered as possible student responses by the person authoring the question. The question in Box 1 is a good example (although is an extreme example in terms of VSA facility): a venous thromboembolism has been confirmed therefore rendering a D-dimer irrelevant, yet just over a quarter of students chose this option in the VSA. More concerning, just over one-third of students would have ordered a CT pulmonary angiogram in a patient with no respiratory symptoms or signs, thereby exposing the patient to a significant dose of unnecessary radiation without any likely therapeutic benefit. It is also possible that further investigation to exclude an occult malignancy would not have been instituted.

VSAs were non-inferior to SBAs on other indices of assessment utility. In terms of feasibility, the electronic delivery platform functioned well and participating medical schools did not report any problems associated with delivering the assessment. The platform also facilitated remote marking. VSAs are more time-consuming to mark than SBAs, but not prohibitively so. The marking time for an individual VSA (and therefore costs) will fall significantly with repeated use as pre-existing marking schemes are re-applied. Furthermore, as students gain experience in the type of answer required, it is possible there will be fewer incorrect

answers to review, which would reduce marking time and costs further. VSAs also had slightly higher internal consistency (a measure of reliability) and comparable question discrimination, as seen in previous small scale pilot studies.[11,12]

This study involved 20 medical schools across the United Kingdom, which were representative of all UK schools in terms of size and location. The large number of medical schools that took part in the study and the overall high number of participants makes this the largest study comparing VSAs with SBAs and suggests that the findings of this study are generalizable across the UK and potentially internationally. Non-completion of the assessments was rare: 1,411 (99.6%) students completed all 50 SBA questions and while more students left blank VSA responses, in terms of evidence of non-completion, only 11 (0.8%) did so for the last question and the maximum number of blank responses for any question (#42) was 24 (1.7%). Previous studies have highlighted the benefits and shortcomings of SBA questions,[2,4-7,13] but our work provides large-scale empirical data to test some of these claims using an alternative and feasible question format as the comparator.

Our study has several limitations. Medical schools agreed to participate and then within each medical school a variable number of students volunteered to participate: therefore some responder bias is likely. Data on participant characteristics were not collected, so whilst we are unable to comment on how representative our sample is in relation to the total final year population of UK medical students, the high number of medical schools and students participating increases the likelihood that our study population is representative. This assessment was formative and was sat at variable timeframes ahead of students' medical school summative assessments (depending on the individual dates for summative assessments which varied for each participating medical school). Students are therefore likely to have prepared and participated in a different way than for a summative assessment, especially as for some schools, final exams were several months after the study date. Students all sat the SBA questions after the VSA questions to ensure there was no cueing in the VSA. This means that positive cue rates may have been biased upwards because

participants had a second look at the questions during the SBA paper, which may have contributed to them arriving at the correct answer along with having the answer options. We did not focus on the negative cue rate (where students answered the VSA correctly and then the SBA incorrectly) in this study. The mean negative cue rate was 3.9%, lower than the 6.1% in a previous study,[13] although our mean was skewed upwards by five questions with negative cue rates in excess of 10% (the median negative cue rate was 2.0%). The negative cue rate was highest on question 27, which asked students to identify the most appropriate test for monitoring respiratory function based on a scenario that described a patient in myasthenic crisis. 48% of students answered correctly in the VSA (choosing forced vital capacity), but 34% of these students then answered the SBA incorrectly, with most being negatively cued by the answer option arterial blood gas.

We have not yet undertaken a criterion-based approach to standard setting using expert judgment, so are unable to determine whether the full cueing effect of SBAs is accounted for in common standard setting processes such as Angoff[15] or Ebel.[16] Furthermore, this study was also not designed to evaluate all components of assessment utility including acceptability to stakeholders. The previous smaller-scale pilots of VSAs reported that students found VSAs more challenging, but appreciated the additional validity they offered.[11,12]

Key extensions to this work should include the study of how SBA and VSA questions are standard set relative to performance and a comparison of the predictive validity of SBA and VSA scores, particularly using measures of performance in clinical settings.

CONCLUSION

VSAs appear to provide a more accurate measure of a candidate's knowledge than SBAs. They also offer greater insight into cognitive errors, thereby offering opportunities to hone teaching, feedback and learning, as well as creating summative assessments with greater validity. Unlike short-answer questions, modified essay formats or clinical reasoning

problems,[9] VSAs are straightforward to deliver in an electronic format and efficient to mark. We need to know that medical students and trainees have the required applied medical knowledge to practice safely without test scores being confounded by the ability to use the cues of SBA answer options. Our results suggest that VSAs could provide a more authentic method of assessing medical knowledge whilst maintaining most of the cost-efficiency of SBAs.

Figure legend

Figure 1: Scatter diagram of VSA facility and the positive cue rate

Legend to quadrants in Figure 1:

Top-left: N=11 questions with low VSA facility (<0.5 or 50%) and a high positive cue rate (>50%).

Top-right: N=7 questions with high VSA facility (>0.5) and a high positive cue rate (>50%).

Bottom-left: N=24 questions with low VSA facility (<0.5) and a low positive cue rate (<50%).

Bottom-right: N=8 questions with high VSA facility (>0.5) and a low positive cue rate (<50%).

Acknowledgements

We would like to thank the Medical Schools Council Assessment Alliance (MSCAA) management and administration team who supported this study. The MSCAA Board provided support and helpful comments on the study protocol. This study required the support of Assessment leads at participating schools, invigilators and students, who all volunteered their time.

Data sharing

The individual item-level data for each student participant are not available.

Transparency declaration

AS (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that there were no discrepancies from the study as originally planned.

Authors' contributions

The study was designed and implemented by AS, RWe, CB and MG. AS and RWe wrote the question paper. AS, RWe and KM undertook the initial marking, which was verified by RWi. CB undertook the data analysis and wrote the first draft of the paper, supported by RWe for the literature review. AS, RWe, MG, KM and RWi provided critical comments on the paper during each round of drafting.

Funding and role of the funder

The MSCAA supported this study through the provision of staff time and travel costs, but no financial payments were made. The MSCAA Board provided suggestions on study design but had no role in data analysis, interpretation and did not input into the writing of this paper, although they did see the draft prior to submission. As noted in the acknowledgements section, MSCAA staff assisted with data collection.

MG is supported by the NIHR Cambridge Biomedical Research Centre.

CB is supported by the NIHR CLAHRC West Midlands initiative. This paper presents independent research and the views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Conflicts of interest

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare the following, in addition to the funding section above:

CB, AS, MG and RW are elected members of the MSCAA Board.

MG and AS are Advisors for the GMC UK Medical Licensing Assessment.

Dissemination

The results of this study have been reported to the participating medical schools. Participating medical students have received feedback on their performance in the assessment. They will have access to the study results on publication of this article.

REFERENCES

1. Coderre SP, Harasym P, Mandin H, Fick G. The impact of two multiple-choice question formats on the problem-solving strategies used by novices and experts. *BMC medical education.* 2004;4(1):23.
2. Heist BS, Gonzalo JD, Durning S, Torre D, Elnicki DM. Exploring clinical reasoning strategies and test-taking behaviors during clinical vignette style multiple-choice examinations: a mixed methods study. *Journal of graduate medical education.* 2014;6(4):709-714.
3. Bibler Zaidi NL, Grob KL, Monrad SM, Kurtz JB, Tai A, Ahmed AZ, et al. Pushing Critical Thinking Skills With Multiple-Choice Questions: Does Bloom's Taxonomy Work? *Academic Medicine.* 2018;93(6):856-859.
4. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet.* 2001;357(9260):945-949.
5. Pugh D, De Champlain A, Touchie C. Plus ÇA Change, plus C'Est Pareil: Making a Continued Case for the Use of MCQs in Medical Education. *Medical teacher.* 2018:1-9.
6. Raduta C. Consequences the extensive use of multiple-choice questions might have on student's reasoning structure. *Rom. Journ. Phys.* 2013;58:1363-1380.
7. McCoubrie P. Improving the fairness of multiple-choice questions: a literature review. *Medical teacher.* 2004;26(8):709-712.
8. Surry LT, Torre D, Durning SJ. Exploring examinee behaviours as validity evidence for multiple-choice question examinations. *Medical education.* 2017;51(10):1075-1085.
9. ten Cate O, Durning SJ. Approaches to Assessing the Clinical Reasoning of Preclinical Students. In: ten Cate O, Custers E, Durning S, eds. *Principles and Practice of Case-based Clinical Reasoning Education*: Springer, Cham; 2018:65-72.

10. Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ.* 1996;1(1):41-67.
11. Sam AH, Field SM, Collares CF, van der Vleuten CP, Wass VJ, Melville C, et al. Very-short-answer questions: reliability, discrimination and acceptability. *Medical education.* 2018;52(4):447-455.
12. Sam AH, Hameed S, Harris J, Meeran K. Validity of very short answer versus single best answer questions for undergraduate assessment. *BMC medical education.* 2016;16(1):266.
13. Schuwirth L, Van der Vleuten C, Donkers H. A closer look at cueing effects in multiple-choice questions. *Medical education.* 1996;30(1):44-49.
14. StataCorp. *Stata Statistical Software: Release 15.* College Station, Texas: LP S; 2016.
15. Angoff W. Scales, norms and equivalent scores. In: Thorndike R, ed. *Educational Measurement*. Washington, DC: American Council on Education; 1971:508-600.
16. Ebel RL. Procedures for the analysis of classroom tests. *Educ Psychol Meas.* 1954;14:352-364.
17. Feldt L. A test of the hypothesis that Cronbach's alpha reliability coefficient is the same for two tests administered to the same sample. *Psychometrika.* 1980;45(1):99-105.
18. Holsgrove G. *Reliability issues in the assessment of small cohorts.* London: General Medical Council; 2010.
19. Curtis L, Burns A. *Unit costs of health and social care 2018.* Canterbury: Kent Uo; 2018.
20. Cohen-Schotanus J, van der Vleuten C. A standard setting method with the best performing students as point of reference: Practical and affordable. *Medical teacher.* 2010;32(2):154-160.

Table 1: Additional data analysis

Component of assessment utility being evaluated	**Method of analysis**
Reliability: internal consistency	Cronbach's alpha coefficient for each type of question compared using the method of Feldt;[17] the Spearman-Brown formula was then used to estimate the number of questions of each type required for an alpha of 0.8.[18]
Cost: time taken to mark VSAs	The total minutes of consultant time required to mark the VSA, costed at the 2016/17 hourly rate for a hospital consultant (including on-costs and overheads) of £108[19] ($143).
Potential educational Impact: effect on pass/fail rates	Cohen standard setting[20] applied to both VSA and SBA total scores; pass/fail decisions for the two assessments were then compared using Cohen's Kappa.
Question discrimination	Pearson correlation coefficient (point-biserial) between students' scores on each question and those on all other questions combined (item-rest correlation) for each type of question; the difference between question types was compared using a Wilcoxon signed rank sum test (for paired, skewed data).

Table 2: Comparison of SBA and VSA questions and scores

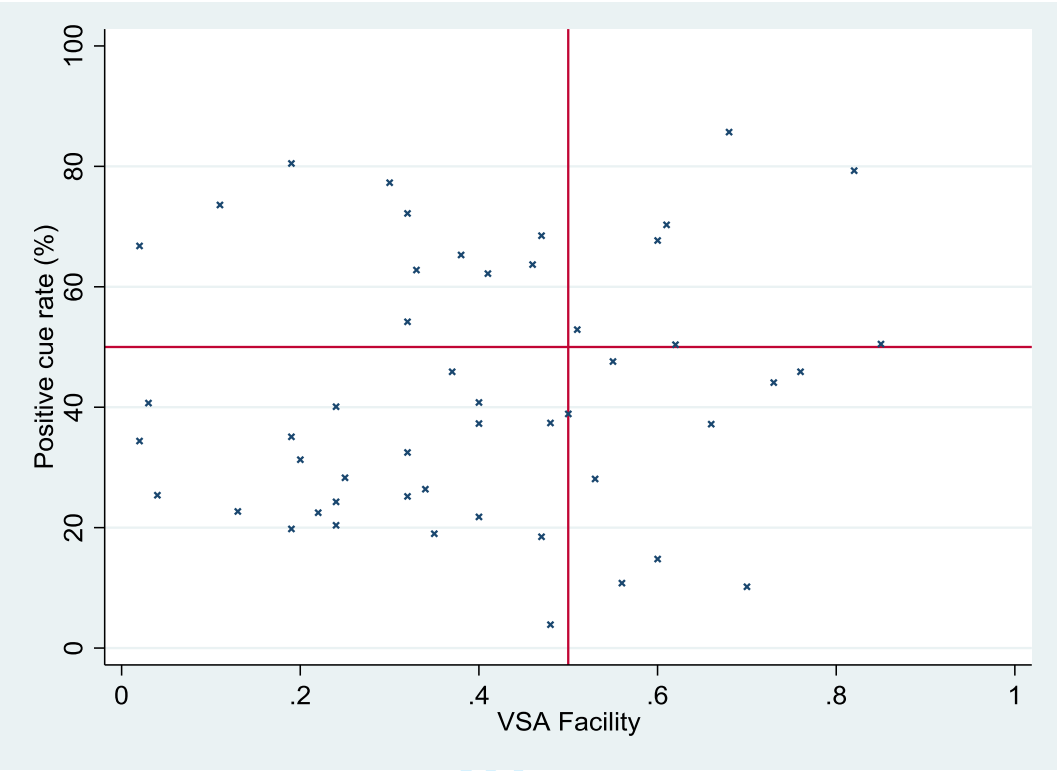
	SBA	VSA	SBA – VSA difference and Statistical significance
Question facility[a] Mean (SD), Range	0.61 (0.20), 0.16 to 0.95	0.40 (0.21), 0.02 to 0.85	0.21 (0.19), -0.32 to 0.65 Paired t-test t=7.89, p<0.001
Positive cue rate (question level) Mean (SD), Range (%)	42.7 (21.3), 3.9 to 85.7		One-sample t-test (NH=<20%) t=7.53, p<0.001
Internal consistency (Cronbach's alpha)	0.693	0.731	-0.038 $F_{1416,1416}$=1.262, p<0.001
Questions required for an alpha of 0.8	89	74	15
Cohen pass mark[b]	28/50	18/50	N/A
Pass rate using Cohen pass mark (%)	71.2	66.3	Kappa = 0.59 z=22.2, p<0.001
Question discrimination Median (IQR), Range	0.184 (0.135 to 0.220), 0.003 to 0.287	0.192 (0.121 to 0.259). -0.006 to 0.395	-0.004 (-0.083 to 0.034), -0.296 to 0.225 Wilcoxon test z=-1.36, p=0.175

[a] Facility: proportion of students answering correctly.

[b] Calculated as 60% of the score of the 95th percentile student and assuming scores due to guessing of 20% for the SBA and 0% for the VSA.

Table 3: Question statistics and themes of questions with VSA facility <0.5 and positive cue rate>50%

Question	SBA Facility	VSA Facility	Difference	Positive cue rate (%)	Theme
9	0.84	0.19	0.65	80.5	Investigations of diabetes insipidus
41	0.82	0.30	0.52	77.3	Diagnosis of cerebellar stroke
3	0.76	0.11	0.65	73.6	Assessment of patient following house fire
25	0.80	0.32	0.48	72.2	Treatment of delirium
16	0.80	0.47	0.34	68.5	Investigations of a neck lump
4	0.68	0.02	0.65	66.8	Further investigation of unprovoked DVT
43	0.78	0.38	0.40	65.3	Determining Glasgow Coma Scale (GCS) Score
13	0.79	0.46	0.32	63.7	Diagnosis of headache
21	0.74	0.33	0.41	62.8	Causative organism of malaria
8	0.76	0.41	0.35	62.2	Diagnosis of (o)esophageal rupture
31	0.66	0.32	0.34	54.2	Management of gout



Appendix 1: Question paper

1. A 73 year old man collapses on the surgical ward 24 hours after having a sigmoid volvulus reduced by sigmoidoscopy.

He has no pulse and an ECG shows asystole. Chest compressions and ventilation are started. The cardiac arrest team are with the patient.

Which is the most appropriate next step in management?

A. Cardiac defibrillation
B. Intravenous 0.9% sodium chloride
C. Intravenous adrenaline/epinephrine
D. Intravenous atropine sulfate
E. Transcutaneous pacing

Correct Answer(s): C

Justification for correct answer

Epinephrine the only recommended treatment for asystole. Atropine no longer recommended.

2. A 78 year old woman has pain in both shoulders, hips and thighs. She is very stiff on waking in the morning and takes 2–3 hours to loosen up. She finds getting dressed difficult.

Investigations:
Erythrocyte sedimentation rate 67 mm/hr (<20)
CRP 87 mg/L (<5)

What is the most appropriate initial treatment?

A. Co-codamol
B. Leflunomide
C. Methotrexate
D. Naproxen
E. Prednisolone

Correct Answer(s): E

Justification for correct answer

The scenario describes typical features of polymyalgia rheumatica, which responds well to prednisolone and poorly to analgesics such as Naproxen or cocodamol. No features to suggest inflammatory arthritis

3. A 36 year old man is rescued from a house fire.

He is alert and talking but has a dull headache. His pulse rate is 98 bpm, BP 139/86 mmHg, respiratory rate 22 breaths per minute and oxygen saturation 100% breathing 15 L/min oxygen via a non-rebreather mask.

Which is the most appropriate parameter to measure?

A. Bicarbonate
B. Carboxyhaemoglobin
C. Haemoglobin
D. Lactate
E. Methaemoglobin

Correct Answer(s): B

Justification for correct answer

Carbon monoxide inhalation likely. Will not show on pulse oximeter as is read as oxyhaemoglobin

4. A 60 year old man has 2 days of a swollen, painful right leg. He has a history of hypertension and takes ramipril. He is otherwise well.

He has a swollen right leg. The remainder of the examination is normal.

Investigations:

Haemoglobin	140 g/L	(130–175)
White cell count	8.0×10^9/L	(3.8–10.0)
Platelets	340×10^9/L	(150–400)
Creatinine	94 µmol/L	(60–120)
Calcium	2.5 mmol/L	(2.2–2.6)
ALT	30 IU/L	(10–50)
ALP	99 IU/L	(25–115)
APTT	30 seconds	(22–41)
PT	12 seconds	(10–12)

Urinalysis: normal

Chest X-ray: normal

Venous duplex ultrasound scan: thrombus in superficial femoral vein

Which is the most appropriate additional investigation?

A. CT of abdomen and pelvis
B. Serum carcinoembryonic antigen
C. Serum prostate specific antigen
D. Serum protein electrophoresis
E. Ultrasonography of abdomen

Correct Answer(s): A

Justification for correct answer

The patient has an unprovoked DVT. Patients should be offered CT scan abdomen and pelvis to help identify possible malignancy.

5. A 65 year old woman had a mechanical aortic valve replacement and coronary revascularisation 3 days ago. She is being treated with dalteparin sodium. She is also taking aspirin long term.

Which is the most appropriate long-term patient management?

A. Apixaban
B. Clopidogrel
C. Continue dalteparin sodium
D. Rivaroxaban
E. Warfarin sodium

Correct Answer(s): E

Justification for correct answer

All patients with mechanical valves require treatment with aspirin and warfarin. Low molecular weight heparin is used as bridging anti-coagulation but not long-term. There is no place for novel oral anticoagulants (yet).

6. The association between maternal smoking during pregnancy and low birthweight can be studied by obtaining smoking histories from pregnant women at the time of first prenatal visit, then assessing birthweight at delivery and analysing the data according to the smoking histories.

Which is the best description of this type of study?

A. Case–control
B. Cross-sectional
C. Prospective cohort
D. Randomised clinical trial
E. Retrospective cohort

Correct Answer(s): C

Justification for correct answer

This is a prospective cohort study where a group of individuals who differ with respect to one or more factors are followed, to determine how these factors affect outcomes. A case-control study finds individuals with a given outcome along with a suitable control group and looks back retrospectively at how many individuals from both groups had the exposure(s) of interest. A cross-sectional study measures prevalence in a population at a given point in time. A randomised clinical trial tests the effects of a specific intervention or interventions against a control group. A retrospective cohort begins after the outcome has occurred and looks back at the exposure of interest.

7. An 18 year old woman has had a chronic skin condition for several years. She has noticed that she gets well-defined areas of scale formation at the sites of minor skin injuries, such as scratches or insect bites, typically when the injuries are healing. The scaling can persist for several weeks or months.

What is the most likely underlying skin condition?

A. Acne vulgaris
B. Eczema
C. Psoriasis
D. Seborrhoeic dermatitis
E. Vitiligo

Correct Answer(s): C

Justification for correct answer

This is typical of Koebnerisation. Psoriasis is by far the commonest underlying cause, though it can also occur in vitiligo.

8. A 64 year old man has vomiting and severe chest pain after eating a large meal.

His temperature is 37.6°C, pulse rate 130 bpm, BP 95/50 mmHg and respiratory rate 30 breaths per minute. There is palpable subcutaneous emphysema on the left side of his neck.

Which is the most likely diagnosis?

A. Diaphragmatic rupture
B. Mallory–Weiss tear
C. Necrotising fasciitis
D. Oesophageal rupture
E. Spontaneous pneumohaemothorax

Correct Answer(s): D

Justification for correct answer

The scenario describes Boerhaave's syndrome (oesophageal rupture).

9. A 54 year old woman has polyuria and the feeling that she needs to drink continuously. She drinks at least 1 litre of water before bedtime and gets up three to four times during the night to pass urine. She has another glass of water each time that she gets up.

Investigations:
Sodium 140 mmol/L (135–146)
Potassium 4.1 mmol/L (3.5–5.3)
Urea 4.5 mmol/L (2.5–7.8)
Creatinine 86 µmol/L (60–120)
Calcium 2.56 mmol/L (2.2–2.6)
Fasting glucose 4.8 mmol/L (3.0–6.0)
Serum osmolality 295 mOsmol/kg (285–295)
Urinary osmolality 86 mOsmol/kg (100–1000)

After 8 hours of a water deprivation test, the serum osmolality is 308 mOsmol/kg and the urinary osmolality is 152 mOsmol/kg.

Following the administration of desmopressin, the serum osmolality is 286 mOsmol/kg and the urinary osmolality is 660 mOsmol/kg.

Which is the most appropriate next investigation?

A. CT scan of thorax, abdomen and pelvis
B. MR scan of pituitary
C. Oral glucose tolerance test
D. Technetium-99 Sestamibi parathyroid scan
E. Supervised fluid restriction and daily weights

Correct Answer(s): B

Justification for correct answer

The test results are consistent with cranial diabetes insipidus.

10. A 30 year old woman has irregular periods, decreased libido and galactorrhoea.

Visual field examination is normal.

Investigations:
Prolactin 5000 mU/L (100–500)
Pregnancy test: negative

MR scan of pituitary shows a 4-mm mass in the sella turcica.

Which is the most appropriate management?

A. Cabergoline
B. Dexamethasone
C. Octreotide
D. Radiotherapy
E. Transsphenoidal surgery

Correct Answer(s): A

Justification for correct answer

Cabergoline (a dopamine agonist) is the first-line treatment for a micro- and macroprolatinomas.#

11. A 23 year old man is admitted to the acute surgical ward with appendicitis and is prepared for theatre. Although he has not eaten for 24 hours, he has been vomiting on and off all day.

Which airway device provides protection for the lungs from regurgitated stomach contents?

A. Guedel (oral) airway
B. i-gel ® (supraglottic) airway
C. Laryngeal mask airway
D. Nasopharyngeal airway
E. Tracheal tube

Correct Answer(s): E

Justification for correct answer

i-gels are often used in cardiac arrest situations as they are easier to place than tracheal tubes. However, only the tracheal tube can seal the trachea off and protect against aspiration.

12. A 32 year old man is referred to a gastroenterology clinic with hepatomegaly. He has a history of type 2 diabetes that is diet controlled. He is taking no regular medication. He is a non-smoker and drinks approximately 16 units of alcohol per week. He is married with no children.

Cardiovascular and respiratory examinations are normal. His abdomen is soft, with a 3 cm palpable liver edge. His BMI is 23 kg/m^2(18–25).

Investigations:

Albumin	38 g/L	(35–50)
ALT	90 IU/L	(10–50)
ALP	112 IU/L	(25–115)
Bilirubin	15 µmol/L	(<17)
Ferritin	710 µg/L	(12–200)
CRP	6 mg/L	(<5)

Which is the most appropriate next investigation?

A. Hepatitis C serology
B. Liver biopsy
C. Reticulocyte count
D. Serum γGT
E. Transferrin saturation

Correct Answer(s): E

Justification for correct answer

Transferrin saturation is the screening test for haemochromatosis.

13. A 30 year old woman has severe headache 24 hours after a spinal anaesthetic.

Her temperature is 37.1°C, pulse rate 90 bpm and BP 120/80 mmHg.

Which is the most likely diagnosis?

A. Low pressure headache
B. Meningitis
C. Migraine
D. Subarachnoid haemorrhage
E. Subdural haemorrhage

Correct Answer(s): A

Justification for correct answer

Most likely cause in this setting - common phenomenon.

14. A 27 year old man is brought to the Emergency Department with left-sided chest pain of sudden onset that is worse on taking a deep breath.

His temperature is 36.8°C, pulse rate 126 bpm, BP 108/60 mmHg, respiratory rate 28 breaths per minute and oxygen saturation 94% breathing air.

Investigations:
ECG: sinus tachycardia

Which is the most appropriate next investigation?

A. Chest X-ray
B. CT pulmonary angiography
C. D dimers
D. Echocardiography
E. Ventilation/perfusion isotope lung scan

Correct Answer(s): A

Justification for correct answer

CXR to rule out other pathology before Well's score and then CTPA (or V/Q) if PE likely or D-dimer if PE unlikely. (NICE CG144) Diagnostic investigations for pulmonary embolism 1.1.7 If a patient presents with signs or symptoms of pulmonary embolism (PE), carry out an assessment of their general medical history, a physical examination and a chest X ray to exclude other causes. [2012]

15. A healthy 23 year old man is scheduled to undergo an elective arthroscopy of his knee. He is to have a general anaesthetic for the operation and asks the pre-operative assessment nurse how long he needs to fast beforehand.

Which are the most appropriate fasting times for clear liquids and solids respectively?

A. 2 h for clear liquids, 2 h for solids
B. 2 h for clear liquids, 6 h for solids
C. 2 h for clear liquids, 12 h for solids
D. 6 h for both clear liquids and solids
E. 6 h for clear liquids, 12 h for solids

Correct Answer(s): B

Justification for correct answer

http://www.aagbi.org/sites/default/files/Perioperative_fasting_in_adults_and_children_.4.pdf Standard national guidelines for elective patients with no problems affecting gastric emptying. Too long a period of fasting is unnecessary whilst residual solid food in the stomach poses a big risk of aspiration/asphyxiation.

16. A 28 year old woman presents to her GP with a neck lump that she noticed incidentally when rubbing her neck.

There is a smooth, non-tender 1.5 cm mobile lump within the thyroid gland.

Investigations:
TSH 2.3 mU/L (0.3–4.2)
Free T4 17 pmol/L (9–25)

Which is the most appropriate next investigation?

A. CT of neck
B. No further investigations
C. Thyroid antibodies
D. Thyroid scintigraphy
E. Ultrasonography of neck

Correct Answer(s): E

Justification for correct answer

This is a non-functional thyroid nodule so needs ultrasound to classify - FNA may then be indicated.

17. A 78 year old man has type 2 diabetes. His clinician does not invite him to join an internet-based self-monitoring programme because she considers him to be too old to engage with it effectively.

What is the most appropriate description of the clinician's approach?

A. Bias
B. Discrimination
C. Inequity
D. Prejudice
E. Stereotyping

Correct Answer(s): B

Justification for correct answer

Discrimination is the unjust or prejudicial treatment of different categories of people.

18. A 47 year old man with hypertension attends for annual review. He takes ramipril (10 mg once daily).

His BP is 138/78 mmHg.

Investigations:

Sodium	139 mmol/L	(135–146)
Potassium	6.2 mmol/L	(3.5–5.3)
Urea	5.0 mmol/L	(2.5–7.8)
Creatinine	90 µmol/L	(60–120)

Which is the most appropriate immediate action?

A. Add indapamide
B. Advise low potassium diet
C. Change ramipril to amlodipine
D. Reduce dose of ramipril
E. Repeat urea and electrolytes

Correct Answer(s): E

Justification for correct answer

This is likely spurious - and needs repeat.

19. A 33 year old woman has 4 months of joint pain and stiffness, predominantly affecting her feet. This is worst in the morning and gradually improves through the day. She feels tired all the time but reports no other health problems.

Which investigation would confirm the most likely diagnosis?

A. Anti-cyclic citrillinated peptide antibody
B. Anti-double-stranded DNA antibodies
C. Antinuclear antibody
D. CRP
E. Rheumatoid factor

Correct Answer(s): A

Justification for correct answer

The correct answer is A (Anti CCP) antibody. This is the test that has the highest specificity for rheumatoid arthritis, which is what the question is asking. CCP antibodies are found in 80% of people with rheumatoid arthritis, but fewer than 0.5% of healthy individuals. Rheumatoid factor is present in up to 10% of the healthy population, and whilst of similar sensitivity to CCP, is much less specific. CRP is a non-specific marker of inflammation, and can often be normal in early rheumatoid. ANA testing is high sensitivity (but low specificity) test for connective tissue disorders such as SLE and Sjorgren's.

20. A 27 year old man has severe central chest pain. He admits to using cocaine shortly before the onset of the chest pain, but says that he had used it on only two previous occasions.

He is distressed and sweating. His pulse rate is 115 bpm and BP 118/68 mmHg. An ECG shows sinus tachycardia with ST elevation in the lateral leads, and several ventricular ectopics.

Which is the mechanism by which cocaine has caused this acute episode?

A. Blockade of myocyte repolarisation
B. Coronary artery spasm
C. Enhanced platelet aggregation
D. Increased systemic vascular resistance
E. Rupture of pre-existing arterial plaques

Correct Answer(s): B

Justification for correct answer

Cocaine-induced ACS.

21. An 18 year old man, who was born in the UK, develops drowsiness and confusion 2 days after returning from visiting his grandparents in Malawi. Over the past week he has had recurrent episodes of high fever.

Investigations:Haemoglobin 92 g/L (130–175)
White cell count 3.2×10^9/L (3.8–10.0)
Platelets 184×10^9/L (150–400)
Blood film parasites visible

Which is the most likely causative organism?

A. Plasmodium falciparum
B. Plasmodium malariae
C. Plasmodium vivax
D. Trypanosoma brucei
E. Trypanosoma cruzi

Correct Answer(s): A

Justification for correct answer

Malaria is the most likely diagnosis. Cerebral involvement makes falciparum more likely. T cruzi is seen in S America. T brucei could cause these symptoms but is much rarer.

22. A 31 year old man visits his GP with a painless lump in his scrotum.

There is a well-defined, non-tender spherical 1 cm mass on the right side of the scrotum. It is superior to the testis and transilluminates.

Which is the most likely diagnosis?

A. Abscess
B. Epididymal cyst
C. Hydrocoele
D. Inguinal hernia
E. Testicular tumour

Correct Answer(s): B

Justification for correct answer

Separate and can get above mass which is cystic is an epididymal cyst (spermatocele) which is benign and usually asymptomatic and managed conservatively. Although a hydrocele trans illuminates it would surround the testis.

23. A 68 year old man has 3 days of worsening vomiting and abdominal pain. He has not passed any stool for 3 days. He has a history of a right hemicolectomy for Dukes' A (T1, N0) bowel cancer 6 months ago.

He is dehydrated and his abdomen is distended.

Which is the most likely diagnosis?

A. Adhesional small bowel obstruction
B. Anastomotic leak
C. Cholecystitis
D. Pancreatitis
E. Tumour recurrence

Correct Answer(s): A

Justification for correct answer

Adhesional bowel obstruction is most likely as he has had a hemicolectomy. Recurrence is less likely because this is Dukes A (early stage). Too late after surgery for anastamotic leak. Pancreatitis and cholecystitis are less likely because they don't cause constipation.

24. A 78 year old patient is admitted with chronic oropharyngeal dysphagia. He has left ventricular systolic dysfunction from ischaemic heart disease. He is breathless on exertion, particularly when climbing stairs. He is being prepared for a percutaneous endoscopic gastrostomy feeding tube. The passage of a nasogastric tube has been unsuccessful, and he is having nil by mouth.

He weighs 70 kg. His pulse rate is 72 bpm and BP 132/80 mmHg.

Which is the correct volume of maintenance fluids (in mL) to prescribe for the next 24 hours?

A. 1750
B. 2450
C. 2800
D. 3000
E. 3250

Correct Answer(s): A

Justification for correct answer

Maintenance fluid requirements for someone with underlying cardiac disease is recommended as 20-25 mL/kg. This gives an upper volume of 25 x 70= 1750. This conforms to current NICE guideline CG 174 (2013).

25. A 75 year old woman is reviewed 4 days after a fractured neck of femur repair. She has been agitated and upset, particularly at night. She has punched nurses and keeps trying to leave the ward. She has seen strange men in black capes entering the ward and believes that they are controlling the hospital. When she was seen in the memory clinic 6 months ago, she was found to have mild cognitive impairment.

What is the most appropriate treatment?

A. Chlordiazepoxide hydrochloride
B. Chlorpromazine hydrochloride
C. Donepezil hydrochloride
D. Haloperidol
E. Memantine hydrochloride

Correct Answer(s): D

Justification for correct answer

This woman has delirium. Therefore, low dose haloperidol would be the best option, according to NICE guidelines.

26. A 72 year old woman has 6 months of increasing pain at the base of her right thumb. She is having difficulty opening jars and sewing. She is otherwise well. No other joints are painful. She is taking regular analgesia.

The first carpometacarpal joint is swollen and tender, with reduced opposition of the thumb.

Which is the most likely diagnosis?

A. De Quervain's tenosynovitis
B. Gout
C. Osteoarthritis
D. Rheumatoid arthritis
E. Septic arthritis

Correct Answer(s): C

Justification for correct answer

This is a classic description of osteoarthritis and a common site Gout and septic arthritis would have a much more acute history. Rheumatoid arthritis would affect multiple joints. The pain of de Quervains tenosynovitis would be felt over the radial aspect of the wrist. Rheumatoid arthritis usually affects multiple joints.

27. A 62 year old man has acute breathlessness with a weak cough, following a recent viral upper respiratory infection. Over the past 4 months, he has had double vision, limb weakness and slurred speech when tired.

His respiratory rate is 18 breaths per minute and oxygen saturation 96% breathing air. He is sweating and using his accessory muscles of inspiration.

Which is the most appropriate test to monitor his respiratory function?

A. Arterial blood gas
B. FEV_1
C. FVC
D. Peak expiratory flow rate
E. Ratio of FEV_1 to FVC

Correct Answer(s): C

Justification for correct answer

Myasthenic crisis is an acute respiratory failure characterised by forced vital capacity (FVC) below 1 L, negative inspiratory force (NIF) of 20 cm H2O or less, and the need for ventilatory support. The use of accessory muscles indicates significant inspiratory weakness. Weak cough indicates weakness of expiratory muscles. Arterial blood gas analysis commonly shows hypercapnia before hypoxia. There should be a low threshold for endotracheal intubation due to rapid deterioration of bulbar and respiratory muscles.

28. A 40 year old man has 4 days of left flank pain associated with fever, nausea and vomiting.

His temperature is 39.6°C, pulse rate 118 bpm and BP 90/40 mmHg. Imaging shows an obstructing proximal left ureteric stone with severe hydronephrosis.

He is treated with intravenous antibiotics and intravenous fluids.

Which is the most appropriate next step in management?

A. Lithotripsy
B. Nephrostomy
C. Retrograde pyelography
D. Ureteric stent
E. Urethral catheter

Correct Answer(s): B

Justification for correct answer

The renal pelvis should be decompressed with a nephrostomy.

29. A 19 year old man has 2 days of right-sided chest pain and breathlessness on exertion. He smokes cannabis and takes cocaine.

His temperature is 36.4°C, pulse rate 108 bpm, BP 112/80 mmHg, respiratory rate 24 breaths per minute and oxygen saturation 94% breathing air. His trachea is central. He has reduced breath sounds at the right apex.

What is the most likely diagnosis?

A. Acute coronary syndrome
B. Coronary artery spasm
C. Pneumonia
D. Pneumothorax
E. Pulmonary embolism

Correct Answer(s): D

Justification for correct answer

Pneumothoraces are more commonly encountered in young, tall men who smoke. There is no past medical history provided in the vignette confirming the presence of existing lung disease - this excludes secondary pneumothorax as a diagnosis.

30. An 80 year old woman is admitted to the Emergency Department after being found collapsed at home. She has central chest pain.

Her pulse rate is 30 bpm, BP 70/40 mmHg and respiratory rate 26 breaths per minute. Her 12-lead ECG shows sinus bradycardia with no evidence of myocardial ischaemia.

Which is the most appropriate initial treatment?

A. Adrenaline/epinephrine
B. Atropine sulfate
C. Dobutamine
D. Normal saline
E. Permanent cardiac pacemaker

Correct Answer(s): B

Justification for correct answer

Atropine is the first line initial treatment for sinus bradycardia.

31. A 62 year old man develops acute pain, redness, swelling and warmth of his right first metatarsophalangeal joint. He has a history of gout and hypertension. His medications are allopurinol, amlodipine and ramipril.

His eGFR is >60 mL/min/1.73 m^2(>60).

Which is the most appropriate next step in his management?

A. Change allopurinol to febuxostat
B. Start naproxen
C. Stop allopurinol
D. Stop amlodipine
E. Stop ramipril

Correct Answer(s): B

Justification for correct answer

The patient has acute gout. The immediate management would be to commence an NSAID.

32. A previously healthy 10 year old boy has deafness of new onset. He has a history of a recent respiratory tract infection.

Tuning fork tests show:
a) when the tuning fork is placed in the middle of his forehead he hears the tone loudest in his right ear;
b) when the tuning fork is held in front of his right external auditory meatus it is quieter than when it is placed on the bone behind the same ear;
c) when the tuning fork is held in front of the left external auditory meatus the sound is louder than when it is placed on the bone behind the same ear.

Which ear(s) is/are affected and which type of hearing loss is this?

A. Bilateral mixed deafness
B. Left conductive deafness
C. Left sensorineural deafness
D. Right conductive deafness
E. Right sensorineural deafness

Correct Answer(s): D

Justification for correct answer

The combination of lateralisation of Weber's test to the right and a negative Rinne's test on the right (i.e. bone conduction louder than air conduction) occurs with right conductive deafness. This also fits with the clinical history of a recent respiratory tract infection.

33. A 55 year old man has 2 days of painful red swelling of his left lower leg. He has a history of type 2 diabetes mellitus and takes metformin.

His temperature is 37.6°C. He has a tender erythematous area extending from the ankle to the proximal calf.

What is the most likely causative organism?

A. Bacteroides species
B. Proteus mirabilis
C. Pseudomonas aeruginosa
D. Staphylococcus epidermidis
E. Streptococcus pyogenes

Correct Answer(s): E

Justification for correct answer

Streptococcus is the most common pathogen in leg cellulitis (including in patients with diabetes).

34. A 55 year old woman has a tender, erythematous, swollen hard cord in the long saphenous vein distribution in her calf. She has a longstanding history of bilateral varicose veins.

An ultrasound scan shows superficial thrombophlebitis without deep vein thrombosis.

Which is the most appropriate treatment?

A. Dipyridamole
B. Flucloxacillin
C. Naproxen
D. Paracetamol
E. Rivaroxaban

Correct Answer(s): C

Justification for correct answer

NSAIDs are the first-line treatment for superficial thrombophlebitis (NICE CKS thrombophlebitis – superficial, May 2017)

35. A 64 year old woman is due to undergo an arthroscopy of her knee. She has type 2 diabetes and takes metformin (500 mg twice daily) and gliclazide (80 mg each morning).

Her glycated haemoglobin is 54 mmol/mol (20–42). She is scheduled first on the morning day case list and is asked to fast from midnight the previous night.

Which is the most appropriate plan for managing her diabetic medication?

A. Continue both drugs and start a variable-rate insulin infusion
B. No change to usual medication
C. Omit both drugs and start a variable-rate insulin infusion
D. Omit gliclazide and continue metformin
E. Omit metformin and continue gliclazide

Correct Answer(s): D

Justification for correct answer

Joint British Diabetes Societies' 2016 guidelines - Management of adults with diabetes undergoing surgery and elective procedures: Improving standards. Principles are to minimise fasting times (hence first on list) and disruption to usual medication. In this setting where the procedure is fairly minor, patient should be eating again at lunchtime and hence there is no call for insulin in any form. Specifics are in Appendix 2 of document. https://www.diabetes.org.uk/resources-s3/2017-09/Surgical%20guidelines%202015%20-%20full%20FINAL%20amended%20Mar%202016_0.pdf

36. A 67 year old man has difficulty walking. He states that he has to raise his left leg higher in the air than normal to avoid scraping his toes on the ground when he walks.

When he raises the left foot from the floor, the ankle assumes a plantar-flexed position with the toes directed towards the floor.

Which nerve is most likely to be affected?

A. Common peroneal
B. Medial plantar
C. Saphenous
D. Superficial peroneal
E. Tibial

Correct Answer(s): A

Justification for correct answer

The patient has foot drop due to loss of active dorsiflexion. The muscles affected are supplied by the common peroneal nerve.

37. A 35 year old woman has 6 months of cyclical pain in both breasts. She has recently noticed a lump in the right breast. There is diffuse nodularity of the axillary tails of both breasts with a discrete 20 mm mass in the upper outer quadrant of the right breast. Fine needle aspiration is performed and 5 mL of brown fluid is removed, with disappearance of the mass. Cytology shows cellular debris with no malignant cells.

What is the most likely diagnosis?

A. Breast abscess
B. Breast carcinoma
C. Fat necrosis
D. Fibroadenoma
E. Fibrocystic disease

Correct Answer(s): E

Justification for correct answer

Fibrocystic disease characteristically causes pain associated with the menstrual cycle. The fine needle aspiration supports this with no malignant cells seen. A fibroadenoma is a solid lump. Pain if present would be localised with fat necrosis. Breast abscess would be a more acute history and again would not be expected to cause bilateral breast pain.

38. A 70 year old man has a brief episode of twitching that starts in his left hand and spreads up the arm over 2 minutes, then stops. His arm feels weak for an hour afterwards. He had an ischaemic stroke affecting his left side 6 months ago with good functional recovery. He has a history of type 2 diabetes mellitus and is taking clopidogrel, metformin, ramipril and simvastatin. He is anxious about a further stroke.

There is no weakness on neurological examination.

Which is the most likely diagnosis?

A. Functional episode
B. Hypoglycaemia
C. Migraine
D. Partial seizure
E. Right hemisphere transient ischaemic attack

Correct Answer(s): D

Justification for correct answer

The description fits with partial seizure affecting his right hemisphere as a result of a previous stroke.

39. A 63 year old woman has 4 months of abdominal bloating, fatigue and nausea.

She is found to have with ovarian cancer. Staging CT is performed to look for lymphatic spread and metastatic disease.

To what regional lymph nodes is her tumour most likely to spread initially?

A. Deep inguinal nodes
B. External iliac nodes
C. Internal iliac nodes
D. Para-aortic nodes
E. Superficial inguinal nodes

Correct Answer(s): D

Justification for correct answer

The main lymphatic drainage of the ovary is to the para-aortic nodes. The iliac nodes are less frequently involved.

40. A 52 year old woman has had three episodes of severe epigastric pain associated with vomiting over the past 3 months. The episodes occurred following eating and lasted for about 1 hour. She has type 2 diabetes mellitus and takes metformin.

Abdominal examination is normal. Her BMI is 35 kg/m^2(18–25).

Investigations:

ALT	15 IU/L	(10–50)
ALP	71 IU/L	(25–115)
Bilirubin	9 µmol/L	(<17)

Ultrasound scan of abdomen: single 2-cm gallstone in gallbladder, common bile duct normal, evidence of fatty liver.

Which is the most appropriate management?

A. Endoscopic retrograde cholangiopancreatography
B. Laparoscopic cholecystectomy
C. MR cholangiopancreatography
D. Open cholecystectomy
E. Ursodeoxycholic acid

Correct Answer(s): B

Justification for correct answer

The patient has symptomatic gallstone disease and laparoscopic cholecystectomy is indicated. Percutaneous cholecystostomy may be used in patients who are not fit for surgery. urosdeoxycholic acid may be used for gallstone dissolution, but is not part of common UK practice.

41. A 56 year old woman develops vertigo, nausea, vomiting and intense occipital headache of sudden onset. She is unable to walk without falling. She has a history of hypertension treated with ramipril.

Her temperature is 37.4°C, pulse rate 94 bpm, BP 146/92 mmHg, respiratory rate 12 breaths per minute and oxygen saturation 96% breathing air. She has multidirectional nystagmus and some clumsiness of her right arm.

Which is the most likely diagnosis?

A. Acute labyrinthitis
B. Benign paroxysmal positional vertigo
C. Cerebellar stroke
D. Ménière's disease
E. Multiple sclerosis

Correct Answer(s): C

Justification for correct answer

Triad of headache, nausea/vomitting and ataxia. Profound imbalance, sudden onset and prominent headache suggest cerebellar stroke.

42. A 79 year old woman has 3 months of a left leg venous ulcer that is slowly healing.

Ankle–brachial pressure Indices are 0.9 on the left side and 1.1 on the right side (0.8–1.2).

Which is the most appropriate management?

A. Compression stockings
B. Diagnostic biopsy
C. Full-length graduated compression bandaging
D. No further management required
E. Repeat ankle–brachial pressure indices in 3 months

Correct Answer(s): A

Justification for correct answer

Despite gradual improvement this venous ulcer would be best managed with stockings, which would also help prevent further lesions.

43. A 26 year old woman sustains a head injury in a motorcycle accident. Her eyes are closed, but she opens them when asked to do. She is confused about what happened and about where she is, but attempts to talk about it. She is repeatedly attempting to remove the cannula from her right wrist.

What is her GCS?

A. 3
B. 7
C. 9
D. 12
E. 14

Correct Answer(s): D

Justification for correct answer

Opens eyes in response to voice = 3 Confused, disoriented = 4 Localizes painful stimuli = 5.

44. A 65 year old man has sudden pain and redness in his right eye. He also has a headache and nausea.

Visual acuity is 6/60 in the right eye. The eye is congested, with a hazy cornea and mid-dilated pupil.

Which is the most likely diagnosis?

A. Acute glaucoma
B. Conjunctivitis
C. Corneal ulcer
D. Scleritis
E. Uveitis

Correct Answer(s): A

Justification for correct answer

All the symptoms and signs described can occur with acute glaucoma. Uveitis whilst causing red eye, headache and visual disturbance is associated with a small pupil. Scleritis, corneal ulcer and conjunctivitis are not generally associated with headache and nausea or a significant drop in visual acuity.

45. A 45 year old woman attends her GP surgery with symptoms of vaginal soreness, itching and discharge. She has had recurrent episodes of vaginal candidiasis over the past 4 months.

The labia minora are red and swollen. A diagnosis of vaginal candidiasis is made.

What is the most appropriate investigation at this stage?

A. Glycated haemoglobin
B. HIV test
C. Sexually transmitted infection screen
D. Test her partner for candidiasis
E. Vaginal pH testing

Correct Answer(s): A

Justification for correct answer

Recurrent candidiasis indicates the need to test for diabetes mellitus.

46. A 68 year old man collapses when rising from a chair and is seen in the emergency department 45 minutes later. He is conscious but has reduced power in his left arm and leg (3/5 and 4/5 respectively) and is slurring his speech. He has a past medical history of COPD and hypertension. He smokes 10 cigarettes per day.

He has bilateral scattered wheeze and carotid bruits on auscultation.

Which is the most appropriate initial radiological investigation?

A. Carotid ultrasonography
B. Cerebral angiography
C. CT cerebral venography
D. CT of head
E. MR imaging of brain

Correct Answer(s): D

Justification for correct answer

This is a probable CVA (presentation and risk factors). Due to rapid presentation he is a candidate for systemic thrombolysis. Non-contrast CT head is the most rapid investigation to exclude intracranial haemorrhage and allow thrombolysis.

47. A 25 year old man sustains 40% full-thickness burns in a house fire. Despite intensive treatment, he becomes breathless and hypotensive. He develops a petechial rash.

His temperature is 38°C, pulse rate 110 bpm, BP 80/50 mmHg and oxygen saturation 96% breathing 40% oxygen.

Investigations:

Haemoglobin	110 g/L	(130–175)
White cell count	4.2×10^9/L	(3.8–10.0)
Platelets	15×10^9/L	(150–400)
APTT	75 seconds	(22–41)
PT	25 seconds	(10–12)
Fibrinogen	0.7 g/dL	(1.5–4.0)
Fibrinogen degradation products	137 mg/mL	(<8)

Which is the most likely diagnosis?

A. Anaphylactic reaction to antibiotics
B. Disseminated intravascular coagulation
C. Fat embolism
D. Immune thrombocytopenic purpura
E. Pulmonary embolism

Correct Answer(s): B

Justification for correct answer

The history, signs and investigations are all most indicative of DIC.

48. A 65 year old woman with newly diagnosed advanced lung cancer has 1 day of breathlessness and 1 week of progressive headache.Her pulse rate is 88 bpm, respiratory rate 20 breaths per minute and oxygen saturation 95% breathing 4 L/min oxygen via nasal prongs. She has a swollen face and neck and distended veins on her chest. Her chest is clear. CT scan of chest shows mediastinal lymphadenopathy compressing the superior vena cava.

Which is the most appropriate initial treatment?

A. Intravenous alteplase
B. Intravenous dexamethasone
C. Intravenous heparin infusion
D. Intravenous mannitol
E. Insert endovenous stent

Correct Answer(s): B

Justification for correct answer

The patient has superior vena cava obstruction. The standard initial treatment is dexamethasone to reduce tumour swelling. There is no evidence of thrombus to justify systemic anticoagulation and systemic thrombolysis has no place. Intravenous mannitol is not used. It is worth noting that although dexamethasone is commonly recommended there is limited controlled data confirming its effectiveness. Insertion of endovenous stent would be considered if there was stridor, but would likely follow intubation and steroids. The following guideline is of use: http://www.palliativecareguidelines.scot.nhs.uk/guidelines/palliative-emergencies/Superior-Vena-Cava-Obstruction.aspx

49. An 80 year old man presents to the Emergency Department with dizziness and melaena of recent onset. He has a metallic mitral valve and is taking aspirin and warfarin sodium. His pulse rate is 80 bpm and BP 122/70 mmHg.

Investigations:

Haemoglobin	105 g/L	(130–175)
White cell count	7.0×10^9/L	(3.8–10.0)
Platelets	676×10^9/L	(150–400)
INR	9.6	(1.0)

He is treated with intravenous vitamin K and is blood cross-matched.

What is the most appropriate additional treatment?

A. Fresh frozen plasma
B. Pantoprazole
C. Protamine sulfate
D. Prothrombin complex concentrate
E. Tranexamic acid

Correct Answer(s): D

Justification for correct answer

Prothrombin complex concentrate is used to reverse warfarin in medical emergencies. It is quicker to administer than FFP and can reverse anti-coagulation within minutes. FFP also carries the risk of allergic reactions, transfusion-related lung injury and volume overload. PCC is therefore considered first-line to reverse warfarin. The other drugs do not reverse warfarin.

50. A 65 year old man has 3 weeks of progressive ankle oedema. He is a lifelong heavy smoker and drinks 12 units of alcohol per week.

His BP is 125/85 mmHg and oxygen saturation 98% breathing air. He has marked bilateral pitting ankle oedema.

Investigations:

Creatinine	85 µmol/L	(60–120)
Urinary protein: creatinine ratio	400 mg/mmol	(<30)
Fasting glucose	5.7 mmol/L	(3.0–6.0)
Total cholesterol	9 mmol/L	(<5.0) Albumin 20 g/L (35–50)

He is treated with furosemide.

Which investigation is most likely to be diagnostic?

A. Chest X-ray
B. Renal arteriography
C. Renal auto-antibody screen
D. Renal biopsy
E. Serum protein electrophoresis

Correct Answer(s): D

Justification for correct answer

Nephrotic syndrome in adults requires renal biopsy to identify the cause, prior to definitive treatment.

Appendix 2: Question-level Statistics

N=1,417 students	Facility		Discrimination (Item-rest correlation)		
Question	SBA	VSA	SBA	VSA	Positive cue rate (%)
1	0.62	0.40	0.22	0.23	40.8
2	0.66	0.50	0.27	0.19	38.9
3	0.76	0.11	0.14	0.26	73.6
4	0.68	0.02	0.03	0.12	66.8
5	0.76	0.66	0.23	0.19	37.2
6	0.76	0.62	0.14	0.18	50.4
7	0.82	0.76	0.16	0.12	45.9
8	0.76	0.41	0.25	0.40	62.2
9	0.84	0.19	0.19	0.20	80.5
10	0.39	0.35	0.13	0.21	19.0
11	0.62	0.53	0.13	0.23	28.1
12	0.42	0.03	0.20	0.11	40.7
13	0.79	0.46	0.21	0.34	63.7
14	0.50	0.56	0.07	0.00	10.8
15	0.53	0.32	0.20	0.19	32.5
16	0.80	0.47	0.16	0.13	68.5
17	0.42	0.19	0.12	0.14	35.1
18	0.35	0.02	0.15	0.00	34.4
19	0.60	0.40	0.18	0.25	37.3
20	0.84	0.60	0.22	0.23	67.7
21	0.74	0.33	0.18	0.23	62.8
22	0.34	0.19	0.19	0.22	19.8
23	0.94	0.68	0.01	0.12	85.7
24	0.40	0.25	0.15	0.11	28.3
25	0.80	0.32	0.13	0.14	72.2

N=1,417 students	Facility		Discrimination (Item-rest correlation)		
Question	SBA	VSA	SBA	VSA	Positive cue rate (%)
26	0.56	0.60	0.15	0.17	14.8
27	0.16	0.48	0.26	0.04	3.9
28	0.38	0.22	0.28	0.26	22.5
29	0.82	0.73	0.18	0.18	44.1
30	0.48	0.32	0.29	0.27	25.2
31	0.66	0.32	0.21	0.14	54.2
32	0.71	0.70	0.26	0.26	10.2
33	0.28	0.04	0.00	-0.01	25.4
34	0.31	0.13	0.11	0.09	22.7
35	0.37	0.24	0.08	0.05	20.4
36	0.76	0.51	0.18	0.30	52.9
37	0.64	0.55	0.20	0.14	47.6
38	0.49	0.34	0.23	0.29	26.4
39	0.49	0.40	0.19	0.31	21.8
40	0.61	0.48	0.18	0.17	37.4
41	0.82	0.30	0.24	0.05	77.3
42	0.30	0.47	0.05	0.35	18.5
43	0.78	0.38	0.13	0.12	65.3
44	0.87	0.61	0.19	0.34	70.3
45	0.41	0.24	0.23	0.16	24.3
46	0.89	0.85	0.20	0.20	50.5
47	0.95	0.82	0.12	0.29	79.3
48	0.53	0.24	0.17	0.25	40.1
49	0.43	0.20	0.20	0.34	31.3
50	0.65	0.37	0.25	0.32	45.9

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2-3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	7-8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	7-8
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	7
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	8-9
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	7-9
Bias	9	Describe any efforts to address potential sources of bias	12-13
Study size	10	Explain how the study size was arrived at	9
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	8-9
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8-9 Table 1
		(*b*) Describe any methods used to examine subgroups and interactions	N/A
		(*c*) Explain how missing data were addressed	9
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	N/A
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	9
		(b) Give reasons for non-participation at each stage	9
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	9
		(b) Indicate number of participants with missing data for each variable of interest	9
Outcome data	15*	Report numbers of outcome events or summary measures	9-11

			Table 2
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	10
		(*b*) Report category boundaries when continuous variables were categorized	N/A
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	10-11
Discussion			
Key results	18	Summarise key results with reference to study objectives	11-12
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	12-13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	14
Generalisability	21	Discuss the generalisability (external validity) of the study results	12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	16

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.