BMJ Open

Diagnostic accuracy study of three alcohol breathalysers marketed for sale to the public

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-005811
Article Type:	Research
Date Submitted by the Author:	30-May-2014
Complete List of Authors:	Ashdown, Helen; University of Oxford, Nuffield Department of Primary Care Health Sciences Fleming, Susannah; University of Oxford, Nuffield Department of Primary Care Health Sciences Spencer, Elizabeth; University of Oxford, Primary Care Health Sciences Thompson, Matthew; University of Oxford, Primary Care Health Sciences; University of Washington, Department of Family Medicine Stevens, Richard; University of Oxford, Department of Primary Health Care
Primary Subject Heading:	Diagnostics
Secondary Subject Heading:	Health policy, Addiction
Keywords:	BIOTECHNOLOGY & BIOINFORMATICS, PUBLIC HEALTH, TOXICOLOGY, PREVENTIVE MEDICINE

Diagnostic accuracy study of three alcohol breathalysers marketed for sale to the public

Helen F Ashdown[1], Susannah Fleming[1], Elizabeth A Spencer[1], Matthew J Thompson[1, 2], Richard J Stevens[1]

[1] Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, United Kingdom

[2] Department of Family Medicine, University of Washington, Box 354696, Seattle, USA

Helen F Ashdown
Academic Clinical Fellow in General Practice

Susannah Fleming
Research Fellow

Elizabeth A Spencer
Nutritional Epidemiologist

Matthew J Thompson
Professor of Family Medicine

Richard J Stevens
University Research Lecturer in Medical Statistics

Correspondence to:
HF Ashdown
helen.ashdown@phc.ox.ac.uk

Word count (main text): 2610

ABSTRACT

Objectives

To assess the diagnostic accuracy of 3 personal breathalyser devices available for sale to the public

Design

Prospective comparative diagnostic accuracy study comparing two single-use breathalysers and one digital multi-use breathalyser (index tests) to a police breathalyser (reference test).

Setting

Establishments licensed to serve alcohol in a UK city

Participants

Of 222 participants recruited, 208 were included in the main analysis. Participants were eligible if they were 18 years old or over, had consumed alcohol and were not intending to drive within the following six hours.

Outcome measures

Sensitivity and specificity of the breathalysers for the detection of being at or over the UK legal driving limit (35 micrograms/100ml breath alcohol concentration).

Results

18% of participants (38/208) were at or over the UK driving limit according to the police breathalyser. The digital multi-use breathalyser had a sensitivity of 89.5% (95% CI 75.9 to 95.8%) and a specificity of 64.1% (95% CI 56.6 to 71.0%). The single-use breathalysers had a sensitivity of 94.7% (95% CI 75.4 to 99.1%) and 26.3% (95% CI 11.8 to 48.8%), and a specificity of 50.6% (95% CI 40.4 to 60.7%) and 97.5% (95% CI 91.4 to 99.3%) respectively. Self-reported alcohol consumption threshold of 5 UK units or fewer had a higher sensitivity than all personal breathalysers.

Conclusions

Sensitivities of alcohol breathalysers marketed to motorists in the UK for driving decisions varied between 26% and 95%, corresponding to approximately three people in four, and one person in twenty, who are falsely reassured when over the limit. Our study strongly suggests that such devices are not fit for purpose, and that policy-makers should reconsider the regulatory frameworks that allow the manufacture and widespread marketing of a medical device with low sensitivity for a decision with potentially catastrophic consequences.

Strengths and limitations of this study

- Personal breathalysers are available for sale to the public in pharmacies for assessing safety to drive following consuming alcohol, some very inexpensively, and in some jurisdictions is now promoted by law.
- Accuracy of personal breathalyser devices has never previously been studied.
- This study tested diagnostic accuracy of these devices in a real-life setting including participant estimation of readings.
- Limitations include the uncontrolled environment of public houses and bars, and wide confidence intervals for some results due to low prevalence of those over the driving limit.
- Even taking the upper confidence limit for sensitivity of the worst performing device (48.8%) would mean that one in two individuals would be falsely reassured and this could lead to potentially dangerous driving decisions.

INTRODUCTION

Road traffic collisions (RTCs) are the second leading cause of death worldwide amongst people aged 5-29 years, estimated at 1.2 million deaths each year, and forecast to rise further by 2020.[1] Consumption of alcohol is an important factor influencing the likelihood and severity of RTCs, and has been found be a causal factor in 17-40% of RTCs worldwide.[2] The risk of an RTC increases rapidly with increasing blood alcohol concentration, with relative risk rising significantly beyond a blood alcohol concentration (BAC) of 0.04g alcohol per 100ml blood (g/dl) to reach a relative risk of approximately 5 at 0.10g/dl.[2] Introduction of maximum legal BAC limits for driving, which vary from 0.02g/dl (e.g. Russia) to 0.15g/dl (e.g. Uganda), has been effective in reducing alcohol-related injuries and deaths.[1]

Alcohol breathalysers, which have long been available to law enforcement agencies, are now marketed direct to consumers, for example in some UK pharmacies and motoring stores. In July 2012, it became a legal requirement for drivers in France to carry a personal breathalyser at all times.[3] Devices marketed to consumers are order of magnitudes lower in cost than those intended for law enforcement: for example at the time of writing the Alcosense Single is more than 300 times cheaper than the Dräger 6510 Home Office certified police breathalyser.[4, 5]

The theoretical accuracy of breath alcohol measurement as a surrogate for blood alcohol measurement has been well-studied,[6] and in the UK breath alcohol forms part of the prescribed legal limit, along with blood and urine alcohol concentration.[7] However to our knowledge the accuracy of devices currently marketed to the consumer/motorist has not been studied. Many such devices carry regulatory approvals such as the Conformité Européenne (CE), French NF certification and British Standard (BSI Kitemark) but in general such marks are statements of engineering quality rather than diagnostic accuracy. We therefore aimed to assess the diagnostic accuracy of personal breathalysers compared to a police breathalyser for detection of being at or over the UK legal driving limit.

METHODS

Index tests and reference standard

We selected for study as index tests the Alcosense Single, the UK counterpart of an NF-approved device widely sold for motorists in France; the Dräger Alco-Check, as a comparable single-use device from a competing manufacturer; and the Alcosense Elite, as an example of a digital multi-use device readily available from pharmacies, high street and online stores (Boots, Halfords, Amazon and others). We selected as reference standard the Dräger Alcotest 6510 device. This has Home Office approval and is standard issue to UK police for use at the roadside,[5] and is approved in the US as an evidential breath testing device (National Highway Traffic Safety Administration Standard 49 FR 48854) meaning readings can be used as evidence for prosecution of drink driving offences. Manufacturer information states measurement precision as ±0.008mg/L or 1.7% of measurement value.[8]

Study participants

We recruited participants from establishments licensed to serve alcohol in the city centre of Oxford, UK, including college bars and public houses. In the absence of prior data with which to estimate sample size, we decided to recruit 200 participants, which would allow us to report the prevalence of intoxication with small standard error (maximum 3.5%). Recruitment took place on evenings during the period from October 2012 to January 2013: 11 study evenings were required to reach a sample size greater than 200. Participants were eligible if they were 18 years or over, had consumed alcohol, and were not intending to drive within the following six hours.

Recruitment and consent

Individuals in these establishments were informally approached by a member of the research team and given preliminary information about the study. Potentially eligible interested participants were then asked to sit with the research team, at tables reserved within the same premises, where they received a full description of the study, were asked to read the study literature, and given an opportunity to ask further questions. Eligibility was checked and written consent taken. Participants were given a card with contact details to use in the event

of withdrawing consent subsequently (for example, the next day): in the event, no participants used this option to withdraw consent.

Study procedures

The research team consisted of between two and four individuals, who followed the breathalyser manufacturers' written instructions in directing their use: a minimum of 20 minutes was enforced between recruitment and using the breathalyser devices, and participants were asked not to drink further alcohol, smoke, use mouthwash, or drink fruit juice during this period. Participants were provided with water and asked to take at least one sip in order to clear any residual alcohol from the upper airway. During this 20 minute period basic demographic details (age and sex) and reported alcoholic drinks consumption during the preceding 12 hours were recorded by the researchers. Participants were not required to remain under observation during the 20 minute period.

Participants then used three breathalysers, at intervals of at least one minute, at the study tables under the supervision of a research team member. Each participant used, in random order, the reference standard, the Alcosense Elite multi-use and a randomly selected single use breathalyser. Randomisation was carried out in advance by study number in random permuted blocks.

A member of the research team recorded digital outputs of the multi-use and reference breathalysers. The single-use breathalysers require subjective assessment of colour of crystals after use: we recorded assessments by both researcher (primary outcome) and participant (secondary outcome). Participants were blinded to the researcher assessment and to the reference result.

Statistical analysis

We calculated sensitivity, specificity, positive and negative predictive values of the breathalysers for the outcome of being at or over the UK legal driving limit[7] of 35 micrograms/100ml or 0.8‰BAC according to the reference standard. Self-reported alcohol consumption, converted to UK alcohol units (1 unit = 8g alcohol) using nutritional tables,[9] was also assessed for diagnostic accuracy of being at or over the UK legal driving limit, and an ROC analysis undertaken to assess different unit thresholds. Statistical calculations were carried out using standard methods[10] and the ROC analysis using Stata (Release 11). Minor

protocol violations were discussed amongst the research team, with sensitivity analyses performed to determine whether there was any difference in overall results due to inclusion or exclusion of these. Participants with missing data (for example sex, self-reported alcohol consumption or participant estimation of result) were included except for analyses involving the missing component itself.

RESULTS

A total of 208 participants were included in the main analysis (see Figure 1 flowchart), of whom 148/207 (71.5%) were male and with a median age of 20. Participants reported having consumed a median of 6 UK units of alcohol (range 1 to 25), equivalent to a median of 46g (range 8 to 204) alcohol.

38 participants (18.3%, 95% CI 11.7 to 27.4%) were at or above the current UK driving limit of 35 μg/100ml according to the reference breathalyser. Table 1 compares performance of the three index breathalysers at detecting those at or over the UK limit. Compared to the reference breathalyser, the Alcosense Elite multi-use breathalyser had a sensitivity of 89.5% (95% CI 75.9 to 95.8%), the Dräger Alco-Check single-use breathalyser had a sensitivity of 94.7% (95% CI 75.4 to 99.1%) and the Alcosense Single breathalyser had a sensitivity of 26.3% (95% CI 11.8 to 48.8%). When analyses were repeated using the participant's interpretation of colour change in the single-use breathalysers instead of researcher's interpretation, sensitivity was 94.7% (95% CI 75.4 to 99.1%) for the Dräger Alco-Check (i.e. identical to researcher estimation) and 16.7% (95% CI 11.8 to 48.8%) for the Alcosense Single (see Appendix 1 supplementary material for full data for participant estimation).

We conducted three sensitivity analyses in turn (i) excluding two results where the participants were suspected to have incorrectly used the device, (ii) a colour-blind participant who may have had difficulty interpreting the colour change of crystals, and (iii) a participant who was suspected to have violated the protocol (consumed alcohol between use of the three devices) showed minimal difference to results overall (see Appendix 1) although the sensitivity of the Dräger Alco-Check increased to 100% in the latter sensitivity analysis. There were no reported adverse events from using any of the breathalyser devices.

Because there is no single standard threshold for driving decisions based on self-estimated alcohol consumption, we calculated sensitivity and specificity for all alcohol unit thresholds (Table 2) and plotted them in ROC space (Figure 2).

DISCUSSION

We have shown that breathalysers available for sale to the public for personal use vary considerably in their performance in detecting being at or over a legal driving limit. Two of the devices tested, the Alcosense Elite digital multi-use breathalyser and Dräger Alco-Check single-use breathalyser had a sensitivity of approximately 90 and 95 per cent respectively. However, even this means that approximately 1 in 20 people over the driving limit would be falsely reassured by these tests. The third device, the Alcosense Single single-use breathalyser, had a sensitivity of only 26 per cent, meaning that only approximately one in four individuals over the legal limit would be identified by this device. Participants (rather than researchers) interpreting results, which is what would occur in real life, reduced sensitivity further to only 17 per cent. This device has a correspondingly high specificity, but specificity is not the safety-critical aspect of performance of a device assessing safety to drive. Surprisingly, we found that self-reported consumption of alcohol was a more sensitive test for being at or over the legal driving limit than the three breathalysers tested for up to 5 UK units of alcohol consumed. None of the devices outperformed simple recall of amount of alcohol consumed up to 5 UK units of alcohol.

Strengths of our study include testing participants in a "real-world" environment including assessment of participants' estimation of breathalyser readings. However, this also brings limitations in that the setting in college bars and public houses could not be rigorously controlled and it is possible that the environment, for example poor lighting or unknown protocol violation through drinking alcohol or smoking immediately prior to testing, could have introduced some inaccuracies. Operating the three breathalysers in close succession and randomisation of order of use of the breathalysers should have reduced the impact of this. As discussed above, we may have underestimated the sensitivity of the Dräger Alco-Check because of a suspected protocol violation. Three participants inadvertently took the breathalysers in a different order to that planned: we decided to include these participants because the aim was to have an overall variety in breathalyser order and this was unlikely to have any impact for such a small number of participants. Participants were blinded to their result on the reference breathalyser until after completing estimation of colour change, however researchers were not and it is possible that this could have introduced some bias. We met our planned recruitment target, but the prevalence of those over the legal limit by the reference standard means that our confidence intervals are fairly wide. Even using the upper

confidence limit for sensitivity, however, the worst-performing breathalyser would still have sensitivity under 50 per cent. Our sample of participants obtained from colleges and pubs may not be representative of the population who purchase personal breathalysers, particularly in age group and quantity of alcohol consumed and socio-demographics; although this might affect sensitivity of self-reported alcohol consumption, it should not impact significantly on the overall accuracy of the breathalyser devices. We tested the index devices on the night of drinking and the generalizability to "morning-after" use may be unclear; blood and breath alcohol concentrations decline with time after drinking alcohol at similar rates and maintain high correlation,[11] but it is possible that diagnostic accuracy of personal breathalysers may differ at lower alcohol concentrations.

We chose the Dräger Alcotest 6510 as our reference standard because it is portable, easy to operate and is the model widely used by the police at the roadside in the UK. It is not currently approved as an evidential device for criminal prosecutions in the UK, although it is in the US, and it is UK Home Office approved for police use at the roadside.[5] An evidential breathalyser would have been better as a reference standard for this reason, but such devices are not portable and would therefore not have been suitable for our study setting. We speculate that devices intended for use by the police might be calibrated to under-read with their margin of error: if this device were under-reading by a small amount, it would have the effect of reducing sensitivity of the personal breathalysers. For convenience we selected index devices easily available in the UK for testing, however the same devices calibrated for different legal limits are sold outside of the UK, and other devices sold elsewhere use similar technology and are similarly priced. Therefore, while we cannot directly apply our results to other countries, we would anticipate similar findings elsewhere.

We have not found other studies testing the accuracy of personal breathalysers. Studies in the US of college student parties have found a mean BAC of 0.077% (standard deviation 0.063%) (for context, the UK driving limit is 0.08%).[12] A Canadian study found a linear relationship between self-reported alcohol consumption and BAC in Emergency Room attendees, up to 7 drinks.[13] However, a US study of college students which compared estimated BAC to measured BAC, found that students tended to over-estimate their levels of consumption when surveyed in the midst of a night of drinking.[14] We did not attempt to convert self-reported alcohol consumption to BAC and only recorded total quantity consumed, which would have had differential effects between individuals dependent on

weight, sex and the time over which the drinks were consumed. However, despite the variation these factors would introduce into BAC, self-reported consumption up to 5 units was still a more sensitive test than the breathalysers tested, and BAC is known to correlate poorly with symptoms of intoxication.[15]

Our research suggests that some personal breathalysers available for sale to the public are not fit for purpose. The use of inaccurate information from breathalysers thought to be accurate could have catastrophic safety implications for drivers. The fact that these devices are sold in well-established pharmacies including national chains does not guarantee sufficient accuracy for safe use. Medical and measurement devices may carry regulatory approvals such as CE- or NF-marking, but this does not appear to correlate with accuracy, and this raises wider questions over how this marking may be perceived by users. A derivative device of the worst-performing breathalyser in our study is widely sold for use in France as part of the new law requiring breathalysers to be carried when driving, and has French NF-approval.[4] Although results from our study cannot be directly applied to the lower French driving limit of 0.05g/dl and a derivative device, it questions the utility of the new law which on the one hand may improve public awareness of drink-driving in general, but also risks ill-informed driving decisions based on inaccurate results from a personal breathalyser. Replication of our results in other settings and with other breathalysers could further inform policy makers planning to introduce similar laws in other jurisdictions, and explore the characteristics of the population who purchase personal breathalysers and how they use the results obtained. Finally, our research raises worrying questions about the level of scrutiny that medical tests intended for sale to the general public undergo in Europe, and raises wider concerns about how diagnostic accuracy in particular is evaluated, and whether any further field evaluations are required for intended users, perceptions of accuracy of such devices and how use of such devices interacts with medical testing in other health care settings.

REFERENCES

1. Peden M, Scurfield R, Sleet D, et al. World report on road traffic injury prevention. Geneva: World Health Organization 2004.
2. Global Road Safety Partnership. Drinking and driving: a road safety manual for decision-makers and practitioners. Geneva: World Health Organization 2007.
3. Légifrance: Le Service Public pour la Diffusion du Droit: Décret n° 2012-284 du 28 février 2012 relatif à la possession obligatoire d'un éthylotest par le conducteur d'un véhicule terrestre à moteur http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000025417826&categorieLien=id [accessed May 2014].
4. Halfords store website. Alcosense Single breathalyser http://www.halfords.com/webapp/wcs/stores/servlet/product_storeId_10001_catalogId_10151_productId_871069_langId_-1_categoryId_226211 [accessed May 2014]
5. UK Breathalysers website. Draeger Police 6510. http://www.ukbreathalysers.com/2_alcohol/test_others/prob/6510.html [accessed May 2014].
6. Gibb KA, Yee AS, Johnston CC, et al. Accuracy and usefulness of a breath alcohol analyzer. *Ann Emerg Med* 1984;13:516-20.
7. Road Traffic Act 1988 Chapter 52 Part I Sections 4-11. Motor vehicles: drink and drugs. Ministry of Justice, UK http://www.legislation.gov.uk/ukpga/1988/52/part/I/crossheading/motor-vehicles-drink-and-drugs [accessed May 2014].
8. Dräger Alcotest 6510 product specification 2012.
9. McCance RA, Widdowson, EM. McCance and Widdowson's The Composition of Foods. 6th edition compiled by the Food Standards Agency and Institute of Food Research, Cambridge, UK: Royal Society of Chemisty 2002.
10. Newcombe RG, Altman, DG. Proportions and their differences. In: Altman DG, Bryant TN, Gardner MJ, eds. Statistics With Confidence. 2nd edition ed. London: BMJ Books 2000:46-54.
11. Jones AW, Andersson L. Comparison of ethanol concentrations in venous blood and end-expired breath during a controlled drinking study. *Forensic Sci Int* 2003;132:18-25.

12. Clapp JD, Min JW, Shillington AM, et al. Person and environment predictors of blood alcohol concentrations: a multi-level study of college parties. *Alcohol Clin Exp Res* 2008;32:100-7.
13. Bond J, Ye Y, Cherpitel CJ, et al. The relationship between self-reported drinking and BAC level in emergency room injury cases: is it a straight line? *Alcohol Clin Exp Res* 2010;34:1118-25.
14. Kraus CL, Salazar NC, Mitchell JR, et al. Inconsistencies between actual and estimated blood alcohol concentrations in a field study of college students: do students really know how much they drink? *Alcohol Clin Exp Res* 2005;29:1672-6.
15. Olson KN, Smith SW, Kloss JS, et al. Relationship between blood alcohol concentration and observable symptoms of intoxication in patients presenting to an emergency department. *Alcohol Alcohol* 2013;48:386-9.

Competing interests
All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements
The authors would like to thank the study participants, and those colleges of the University of Oxford and public houses in Oxford which permitted them to recruit on their premises and Michael Ashdown for assistance with legal literature.

Contributors
All authors contributed to study design, data collection, data interpretation and helped write the paper. RJS conceived the idea for the study. HFA, EAS and RJS analysed the data. HFA is guarantor. All authors approved the final version of the manuscript.

Ethical approval
The study was approved by the University of Oxford Medical Sciences Division Interdivisional Research Ethics Committee (reference MSD/IDREC/2011/13). All participants gave informed consent before taking part. No remuneration or incentives were provided to participants for taking part.

Funding
This paper summarises independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0407-10347). The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Sponsorship
The University of Oxford acted as study sponsor and had no role in study design, data collection, analysis or interpretation, writing of the paper or the decision to submit for publication. All authors were independent from funders and sponsors.

Data sharing

All authors had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. Full data are available from the corresponding author on request. Consent for data sharing was not obtained but the presented data are anonymised and risk of identification is low.

Transparency declaration

The lead author and the manuscript's guarantor affirm 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 any discrepancies from the study as planned have been explained.

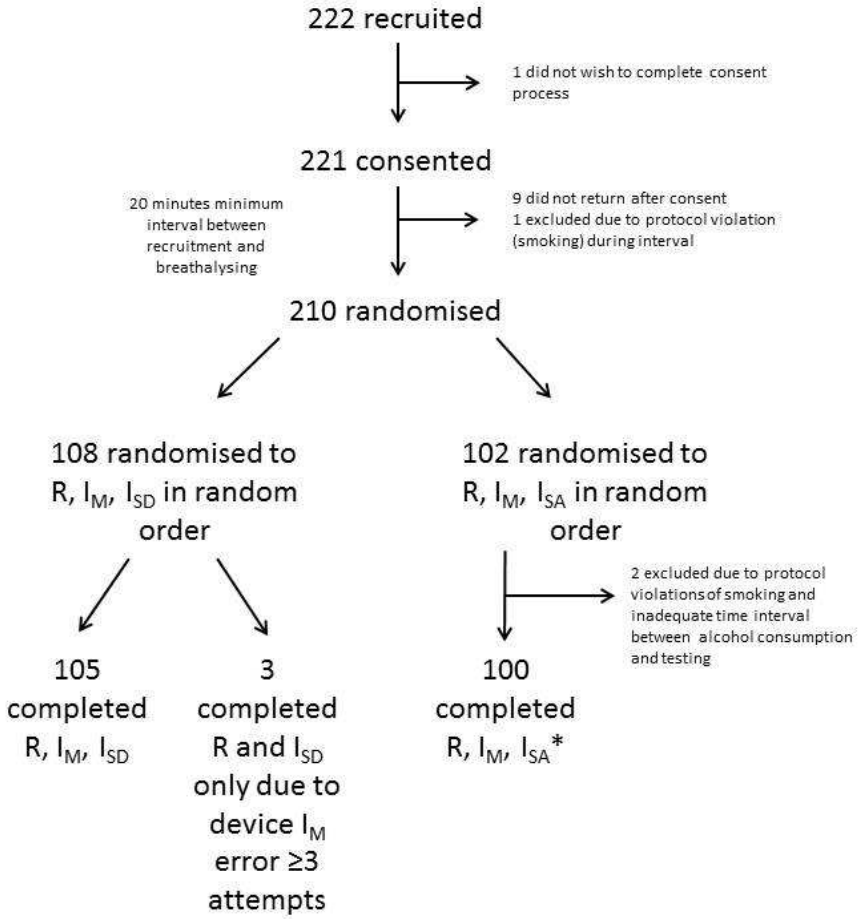


Figure 1: Study flow chart. All participants (except exclusions detailed in figure) were tested with the Dräger Alcotest 6510 (R), the Alcosense Elite (I_M), and one of either the Dräger Alco-Check (I_{SD}) or the Alcosense Single (I_{SA}). The order of undertaking each breathalyser, and the selection of I_{SD} or I_{SA} was determined by randomisation. *3 participants left before analysis therefore results for participant estimation of I_{SA} are only available for 97 participants.

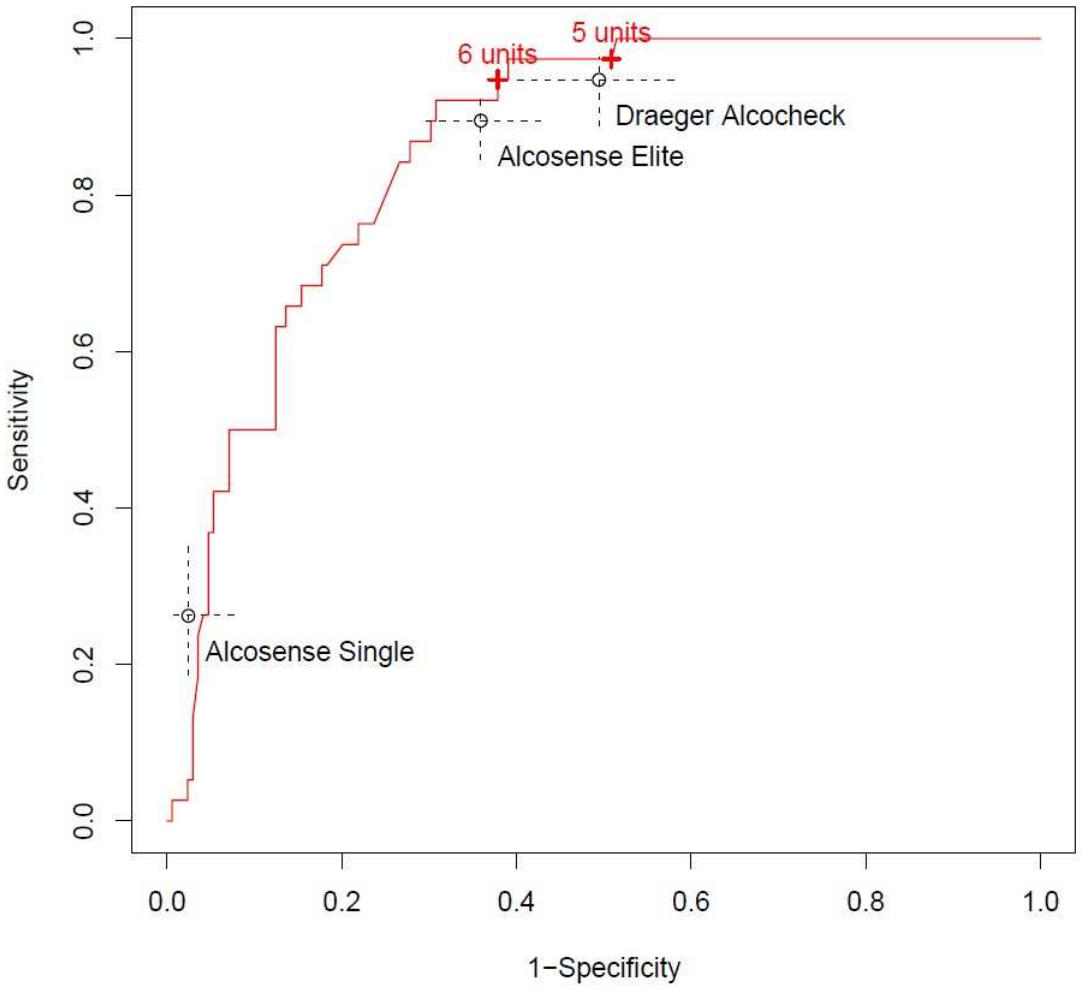


Figure 2: ROC curve of self-reported alcohol consumption in UK units, with comparative sensitivity and specificity for breathalysers tested.

Breathalyser	Total participants	Correctly positive by reference standard	Correctly negative by reference standard	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)
Alcosense Elite multi-use	205	34/38	107/167	89.5 (75.9 to 95.8)	64.1 (56.6 to 71.0)	36.2 (27.2 to 46.3)	96.4 (91.1 to 98.6)
Dräger Alco-Check single-use	108	18/19	45/89	94.7 (75.4 to 99.1)	50.6 (40.4 to 60.7)	29.0 (19.2 to 41.3)	97.8 (88.7 to 99.6)
Alcosense Single single-use	100	5/19	79/81	26.3 (11.8 to 48.8)	97.5 (91.4 to 99.3)	71.4 (35.9 to 91.8)	85.0 (76.3 to 90.8)

Table 1: Diagnostic accuracy of index breathalysers compared to reference police breathalyser, using researcher interpretation of single-use breathalysers. 95% confidence intervals shown in parentheses.

Threshold of self-reported alcohol consumption (UK units)	Total participants	Correctly positive by reference standard	Correctly negative by reference standard	Sensitivity (%)	Specificity (%)
1	207	38/38	9/169	100.0 (90.8 to 100.0)	5.3 (2.8 to 9.8)
2	207	38/38	23/169	100.0 (90.8 to 100.0)	13.6 (9.2 to 19.6)
3	207	38/38	52/169	100.0 (90.8 to 100.0)	30.8 (24.3 to 38.1)
4	207	38/38	71/169	100.0 (90.8 to 100.0)	42.0 (34.8 to 49.5)
5	207	37/38	84/169	97.4 (86.5 to 99.5)	49.7 (42.3 to 57.2)
6	207	35/38	105/169	92.1 (79.2 to 97.3)	62.1 (54.6 to 69.1)
7	207	34/38	117/169	89.5 (75.9 to 95.8)	69.2 (61.9 to 75.7)
8	207	29/38	130/169	76.3 (60.8 to 87.0)	76.9 (70.0 to 82.6)
9	207	26/38	142/169	68.4 (52.5 to 80.9)	84.0 (77.7 to 88.8)
10	207	20/38	148/169	52.6 (37.3 to 67.5)	87.6 (81.8 to 91.7)

Table 2: Diagnostic accuracy of self-reported alcohol consumption compared to reference police breathalyser. 1 UK unit is equivalent to 10ml or 8g alcohol (N.B. total 207 partcipants due to missing data on alcohol consumption for one participant). 95% confidence intervals shown in parentheses.

Web appendix: full 2x2 table data and diagnostic accuracy results for researcher and participant estimation and sensitivity analyses

Main analysis

Alcosense Elite

		Reference		
		Positive	Negative	Total
Index	Positive	34	60	94
	Negative	4	107	111
	Total	38	167	205

		Lower limit CI	*Upper limit CI*
Sensitivity	89.47%	*75.87%*	*95.83%*
Specificity	64.07%	*56.55%*	*70.96%*
PPV	36.17%	*27.18%*	*46.25%*
NPV	96.40%	*91.10%*	*98.59%*

Dräger Alco-Check

		Reference		
		Positive	Negative	Total
Index	Positive	18	44	62
	Negative	1	45	46
	Total	19	89	108

		Lower limit CI	*Upper limit CI*
Sensitivity	94.74%	*75.36%*	*99.06%*
Specificity	50.56%	*40.37%*	*60.71%*
PPV	29.03%	*19.22%*	*41.29%*
NPV	97.83%	*88.66%*	*99.62%*

Alcosense Single

		Reference		
		Positive	Negative	Total
Index	Positive	5	2	7
	Negative	14	79	93
	Total	19	81	100

		Lower limit CI	*Upper limit CI*
Sensitivity	26.32%	*11.81%*	*48.79%*
Specificity	97.53%	*91.44%*	*99.32%*
PPV	71.43%	*35.89%*	*91.78%*
NPV	84.95%	*76.30%*	*90.82%*

Participant estimation

Alcosense Elite

Not applicable as no participant estimation
(i.e. same as main analysis)

Dräger Alco-Check

		Reference		
		Positive	Negative	Total
Index	Positive	18	42	60
	Negative	1	47	48
	Total	19	89	108

		Lower limit CI	*Upper limit CI*
Sensitivity	94.74%	*75.36%*	*99.06%*
Specificity	52.81%	*42.54%*	*62.85%*
PPV	30.00%	*19.90%*	*42.51%*
NPV	97.92%	*89.10%*	*99.63%*

Alcosense Single

		Reference		
		Positive	Negative	Total
Index	Positive	3	4	7
	Negative	15	75	90
	Total	18	79	97

		Lower limit CI	*Upper limit CI*
Sensitivity	16.67%	*11.81%*	*48.79%*
Specificity	94.94%	*91.44%*	*99.32%*
PPV	42.86%	*35.89%*	*91.78%*
NPV	83.33%	*76.30%*	*90.82%*

Participant estimation - sensitivity analysis excluding colour blind person (A/168)

Alcosense Elite

Not applicable as no participant estimation
(i.e. same as main analysis)

Dräger Alco-Check

		Reference		
		Positive	Negative	Total
Index	Positive	18	41	59
	Negative	1	47	48
	Total	19	88	107

		Lower limit CI	*Upper limit CI*
Sensitivity	94.74%	*75.36%*	*99.06%*
Specificity	53.41%	*43.06%*	*63.47%*
PPV	30.51%	*20.25%*	*43.15%*
NPV	97.92%	*89.10%*	*99.63%*

Alcosense Single

Not applicable as colour blind person not in this group
(i.e. same as participant estimation above)

Researcher estimation - sensitivity analysis excluding suspect protocol violation (A/111)

Alcosense Elite

		Reference		
		Positive	Negative	Total
Index	Positive	34	60	94
	Negative	3	107	110
	Total	37	167	204

		Lower limit CI	*Upper limit CI*
Sensitivity	91.89%	*78.70%*	*97.20%*
Specificity	64.07%	*56.55%*	*70.96%*
PPV	36.17%	*27.18%*	*46.25%*
NPV	97.27%	*92.29%*	*99.07%*

Dräger Alco-Check

Not applicable as protocol violation person not in this group
(i.e. same as main analysis)

Alcosense Single

		Reference		
		Positive	Negative	Total
Index	Positive	5	2	7
	Negative	13	79	92
	Total	18	81	99

		Lower limit CI	*Upper limit CI*
Sensitivity	27.78%	*12.50%*	*50.87%*
Specificity	97.53%	*91.44%*	*99.32%*
PPV	71.43%	*35.89%*	*91.78%*
NPV	85.87%	*77.31%*	*91.55%*

Researcher estimation - sensitivity analysis excluding potential faults with devices (A/165 and A/179)

Alcosense Elite

Not applicable as no potential errors with Alcosense Elite
(i.e. same as main analysis)

Dräger Alco-Check

		Reference		
		Positive	Negative	Total
Index	Positive	18	44	62
	Negative	0	45	45
	Total	18	89	107

		Lower limit CI	*Upper limit CI*
Sensitivity	100.00%	*82.41%*	*100.00%*
Specificity	50.56%	*40.37%*	*60.71%*
PPV	29.03%	*19.22%*	*41.29%*
NPV	100.00%	*92.13%*	*100.00%*

Alcosense Single

		Reference		
		Positive	Negative	Total
Index	Positive	5	2	7
	Negative	14	78	92
	Total	19	80	99

		Lower limit CI	*Upper limit CI*
Sensitivity	26.32%	*11.81%*	*48.79%*
Specificity	97.50%	*91.34%*	*99.31%*
PPV	71.43%	*35.89%*	*91.78%*
NPV	84.78%	*76.06%*	*90.71%*

Diagnostic accuracy Study of three alcohol breathalyzers marketed for sale to the public

Aghdam HF et al

Table 1. STARD checklist for the reporting of studies of diagnostic accuracy.

Section and Topic	Item #		On page #
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	1
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	4
METHODS		Describe	
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where the data were collected.	5
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	5
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4? If not, specify how participants were further selected.	5
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	5
Test methods	7	The reference standard and its rationale.	
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	5
	9	Definition of and rationale for the units, cutoffs and/or categories of the results of the index tests and the reference standard.	6
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard.	5
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	5/6
Statistical methods	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	6
	13	Methods for calculating test reproducibility, if done.	N/a
RESULTS		Report	
Participants	14	When study was done, including beginning and ending dates of recruitment.	5
	15	Clinical and demographic characteristics of the study population (e.g. age, sex, spectrum of presenting symptoms, comorbidity, current treatments, recruitment centers).	7
	16	The number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test (a flow diagram is strongly recommended).	7/ Figure 1
Test results	17	Time interval from the index tests to the reference standard, and any treatment administered between.	6
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	7
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	7/ Table 1/ Web appendix
	20	Any adverse events from performing the index tests or the reference standard.	7
Estimates	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	7/Table 1/ Web appendix
	22	How indeterminate results, missing responses and outliers of the index tests were handled.	6/7/ Web appendix
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	7/ Web appendix
	24	Estimates of test reproducibility, if done.	N/a
DISCUSSION	25	Discuss the clinical applicability of the study findings.	8-10

BMJ Open

Diagnostic accuracy study of three alcohol breathalysers marketed for sale to the public

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-005811.R1
Article Type:	Research
Date Submitted by the Author:	10-Sep-2014
Complete List of Authors:	Ashdown, Helen; University of Oxford, Nuffield Department of Primary Care Health Sciences Fleming, Susannah; University of Oxford, Nuffield Department of Primary Care Health Sciences Spencer, Elizabeth; University of Oxford, Primary Care Health Sciences Thompson, Matthew; University of Oxford, Primary Care Health Sciences; University of Washington, Department of Family Medicine Stevens, Richard; University of Oxford, Department of Primary Health Care
Primary Subject Heading:	Diagnostics
Secondary Subject Heading:	Health policy, Addiction
Keywords:	BIOTECHNOLOGY & BIOINFORMATICS, PUBLIC HEALTH, TOXICOLOGY, PREVENTIVE MEDICINE

Diagnostic accuracy study of three alcohol breathalysers marketed for sale to the public

Helen F Ashdown[1], Susannah Fleming[1], Elizabeth A Spencer[1], Matthew J Thompson[1, 2], Richard J Stevens[1]

[1] Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, United Kingdom

[2] Department of Family Medicine, University of Washington, Box 354696, Seattle, USA

Helen F Ashdown
Clinical Researcher in General Practice

Susannah Fleming
Research Fellow

Elizabeth A Spencer
Nutritional Epidemiologist

Matthew J Thompson
Professor of Family Medicine

Richard J Stevens
University Research Lecturer in Medical Statistics

Correspondence to:
HF Ashdown
helen.ashdown@phc.ox.ac.uk

Word count (main text): 2557

ABSTRACT

Objectives

To assess the diagnostic accuracy of 3 personal breathalyser devices available for sale to the public

Design

Prospective comparative diagnostic accuracy study comparing two single-use breathalysers and one digital multi-use breathalyser (index tests) to a police breathalyser (reference test).

Setting

Establishments licensed to serve alcohol in a UK city

Participants

Of 222 participants recruited, 208 were included in the main analysis. Participants were eligible if they were 18 years old or over, had consumed alcohol and were not intending to drive within the following six hours.

Outcome measures

Sensitivity and specificity of the breathalysers for the detection of being at or over the UK legal driving limit (35 micrograms/100ml breath alcohol concentration).

Results

18% of participants (38/208) were at or over the UK driving limit according to the police breathalyser. The digital multi-use breathalyser had a sensitivity of 89.5% (95% CI 75.9 to 95.8%) and a specificity of 64.1% (95% CI 56.6 to 71.0%). The single-use breathalysers had a sensitivity of 94.7% (95% CI 75.4 to 99.1%) and 26.3% (95% CI 11.8 to 48.8%), and a specificity of 50.6% (95% CI 40.4 to 60.7%) and 97.5% (95% CI 91.4 to 99.3%) respectively. Self-reported alcohol consumption threshold of 5 UK units or fewer had a higher sensitivity than all personal breathalysers.

Conclusions

Sensitivities of alcohol breathalysers marketed to motorists in the UK for driving decisions varied between 26% and 95%, corresponding to approximately three people in four, and one person in twenty, who are falsely reassured when over the limit. Our study strongly suggests that such devices are not fit for purpose, and that policy-makers should reconsider the regulatory frameworks that allow the manufacture and widespread marketing of a medical device with low sensitivity for a decision with potentially catastrophic consequences.

Strengths and limitations of this study

- Personal breathalysers are available for sale to the public in pharmacies for assessing safety to drive following consuming alcohol, some very inexpensively, and in some jurisdictions is now promoted by law.
- Accuracy of personal breathalyser devices has never previously been studied.
- This study tested diagnostic accuracy of these devices in a real-life setting including participant estimation of readings.
- Limitations include the uncontrolled environment of public houses and bars, use of a pragmatic reference standard, and wide confidence intervals for some results due to low prevalence of those over the driving limit.
- However, our conclusions for the worst performing device are robust even against this uncertainty, since even the upper confidence limit for sensitivity of the worst performing device (48.8%) would still mean that about one in two individuals would be falsely reassured and this could lead to potentially dangerous driving decisions.

INTRODUCTION

Road traffic collisions (RTCs) are the second leading cause of death worldwide amongst people aged 5-29 years, estimated at 1.2 million deaths each year, and forecast to rise further by 2020.[1] Consumption of alcohol is an important factor influencing the likelihood and severity of RTCs, and has been found be a causal factor in 17-40% of RTCs worldwide.[2] The risk of an RTC increases rapidly with increasing blood alcohol concentration, with relative risk rising significantly beyond a blood alcohol concentration (BAC) of 0.04g alcohol per 100ml blood (g/dl) to reach a relative risk of approximately 5 at 0.10g/dl.[2] Introduction of maximum legal BAC limits for driving, which vary from 0.02g/dl (e.g. Russia) to 0.15g/dl (e.g. Uganda), has been effective in reducing alcohol-related injuries and deaths.[1]

Alcohol breathalysers, which have long been available to law enforcement agencies, are now marketed direct to consumers, for example in some UK pharmacies and motoring stores. In July 2012, it became a legal requirement for drivers in France to carry a personal breathalyser at all times.[3] Devices marketed to consumers are order of magnitudes lower in cost than those intended for law enforcement: for example at the time of writing the Alcosense Single is more than 300 times cheaper than the Dräger 6510 Home Office certified police breathalyser.[4, 5]

The theoretical accuracy of breath alcohol measurement as a surrogate for blood alcohol measurement has been well-studied,[6] and in the UK breath alcohol forms part of the prescribed legal limit, along with blood and urine alcohol concentration.[7] However to our knowledge the accuracy of devices currently marketed to the consumer/motorist has not been studied. Many such devices carry regulatory approvals such as the Conformité Européenne (CE), French NF certification and British Standard (BSI Kitemark) but in general such marks are statements of engineering quality rather than diagnostic accuracy. We therefore aimed to assess the diagnostic accuracy of personal breathalysers compared to a police breathalyser for detection of being at or over the UK legal driving limit.

METHODS

Index tests and reference standard

We selected for study as index tests the Alcosense Single, the UK counterpart of an NF-approved device widely sold for motorists in France; the Dräger Alco-Check, as a comparable single-use device from a competing manufacturer; and the Alcosense Elite, as an example of a digital multi-use device readily available from pharmacies, high street and online stores (Boots, Halfords, Amazon and others). We selected as reference standard the Dräger Alcotest 6510 device. This has Home Office approval and is standard issue to UK police for use at the roadside,[5] and is approved in the US as an evidential breath testing device (National Highway Traffic Safety Administration Standard 49 FR 48854) meaning readings can be used as evidence for prosecution of drink driving offences. Manufacturer information states measurement precision as ±0.008mg/L or 1.7% of measurement value.[8]

Study participants

We recruited participants from establishments licensed to serve alcohol in the city centre of Oxford, UK, including college bars and public houses. In the absence of prior data with which to estimate sample size, we decided to recruit 200 participants, which would allow us to report the prevalence of intoxication with small standard error (maximum 3.5%). Recruitment took place on evenings during the period from October 2012 to January 2013: 11 study evenings were required to reach a sample size greater than 200. Participants were eligible if they were 18 years or over, had consumed alcohol, and were not intending to drive within the following six hours.

Recruitment and consent

Individuals in these establishments were informally approached by a member of the research team and given preliminary information about the study. Potentially eligible interested participants were then asked to sit with the research team, at tables reserved within the same premises, where they received a full description of the study, were asked to read the study literature, and given an opportunity to ask further questions. Eligibility was checked and written consent taken. Participants were given a card with contact details to use in the event of withdrawing consent subsequently (for example, the next day): in the event, no participants used this option to withdraw consent.

Study procedures

The research team consisted of between two and four individuals, who followed the breathalyser manufacturers' written instructions in directing their use: a minimum of 20 minutes was enforced between recruitment and using the breathalyser devices, and participants were asked not to drink further alcohol, smoke, use mouthwash, or drink fruit juice during this period. Participants were provided with water and asked to take at least one sip in order to clear any residual alcohol from the upper airway. During this 20 minute period basic demographic details (age and sex) and reported alcoholic drinks consumption during the preceding 12 hours were recorded by the researchers. Participants were not required to remain under observation during the 20 minute period.

Participants then used three breathalysers, at intervals of at least one minute, at the study tables under the supervision of a research team member. Each participant used, in random order, the reference standard, the Alcosense Elite multi-use and a randomly selected single use breathalyser. Randomisation was carried out in advance by study number in random permuted blocks.

A member of the research team recorded digital outputs of the multi-use and reference breathalysers. The single-use breathalysers require subjective assessment of colour of crystals after use: we recorded assessments by both researcher (primary outcome) and participant (secondary outcome). Participants were blinded to the researcher assessment and to the reference result.

Statistical analysis

We calculated sensitivity, specificity, positive and negative predictive values of the breathalysers for the outcome of being at or over the UK legal driving limit[7] of 35 micrograms/100ml or 0.8‰BAC according to the reference standard. Self-reported alcohol consumption, converted to UK alcohol units (1 unit = 8g alcohol) using nutritional tables,[9] was also assessed for diagnostic accuracy of being at or over the UK legal driving limit, and an ROC analysis undertaken to assess different unit thresholds. Statistical calculations were carried out using standard methods[10] and the ROC analysis using Stata (Release 11). Minor protocol violations were discussed amongst the research team, with sensitivity analyses performed to determine whether there was any difference in overall results due to inclusion or

exclusion of these. Participants with missing data (for example sex, self-reported alcohol consumption or participant estimation of result) were included except for analyses involving the missing component itself.

RESULTS

A total of 208 participants were included in the main analysis (see Figure 1 flowchart), of whom 148/207 (71.5%) were male and with a median age of 20. Participants reported having consumed a median of 6 UK units of alcohol (range 1 to 25), equivalent to a median of 46g (range 8 to 204) alcohol.

38 participants (18.3%, 95% CI 11.7 to 27.4%) were at or above the current UK driving limit of 35 μg/100ml according to the reference breathalyser. Table 1 compares performance of the three index breathalysers at detecting those at or over the UK limit. Compared to the reference breathalyser, the Alcosense Elite multi-use breathalyser had a sensitivity of 89.5% (95% CI 75.9 to 95.8%), the Dräger Alco-Check single-use breathalyser had a sensitivity of 94.7% (95% CI 75.4 to 99.1%) and the Alcosense Single breathalyser had a sensitivity of 26.3% (95% CI 11.8 to 48.8%). When analyses were repeated using the participant's interpretation of colour change in the single-use breathalysers instead of researcher's interpretation, sensitivity was 94.7% (95% CI 75.4 to 99.1%) for the Dräger Alco-Check (i.e. identical to researcher estimation) and 16.7% (95% CI 11.8 to 48.8%) for the Alcosense Single (see Appendix 1 supplementary material for full data for participant estimation).

We conducted three sensitivity analyses in turn (i) excluding two results where the participants were suspected to have incorrectly used the device, (ii) a colour-blind participant who may have had difficulty interpreting the colour change of crystals, and (iii) a participant who was suspected to have violated the protocol (consumed alcohol between use of the three devices) showed minimal difference to results overall (see Appendix 1) although the sensitivity of the Dräger Alco-Check increased to 100% in the latter sensitivity analysis. There were no reported adverse events from using any of the breathalyser devices.

Because there is no single standard threshold for driving decisions based on self-estimated alcohol consumption, we calculated sensitivity and specificity for all alcohol unit thresholds (Table 2) and plotted them in ROC space (Figure 2).

DISCUSSION

We have shown that breathalysers available for sale to the public for personal use vary considerably in their performance in detecting being at or over a legal driving limit. Two of the devices tested, the Alcosense Elite digital multi-use breathalyser and Dräger Alco-Check single-use breathalyser had a sensitivity of approximately 90 and 95 per cent respectively. However, even this means that approximately 1 in 20 people over the driving limit would be falsely reassured by these tests. The third device, the Alcosense Single single-use breathalyser, had a sensitivity of only 26 per cent, meaning that only approximately one in four individuals over the legal limit would be identified by this device. Participants (rather than researchers) interpreting results, which is what would occur in real life, reduced sensitivity further to only 17 per cent. This device has a correspondingly high specificity, but specificity is not the safety-critical aspect of performance of a device assessing safety to drive. Surprisingly, we found that self-reported consumption of alcohol was a more sensitive test for being at or over the legal driving limit than the three breathalysers tested for up to 5 UK units of alcohol consumed. None of the devices outperformed simple recall of amount of alcohol consumed up to 5 UK units of alcohol.

Strengths of our study include testing participants in a "real-world" environment including assessment of participants' estimation of breathalyser readings, which enables generalizability to everyday life. However, this pragmatic approach also brings limitations in that the setting in college bars and public houses could not be rigorously controlled and it is possible that the environment, for example poor lighting or unknown protocol violation through drinking alcohol or smoking immediately prior to testing, could have introduced some inaccuracies. Operating the three breathalysers in close succession and randomisation of order of use of the breathalysers should have reduced the impact of this. As discussed above, we may have underestimated the sensitivity of the Dräger Alco-Check because of a suspected protocol violation. Three participants inadvertently took the breathalysers in a different order to that planned: we decided to include these participants because the aim was to have an overall variety in breathalyser order and this was unlikely to have any impact for such a small number of participants. Participants were blinded to their result on the reference breathalyser until after completing estimation of colour change, however researchers were not and it is possible that this could have introduced some bias. We met our planned recruitment

target, but the prevalence of those over the legal limit by the reference standard means that our confidence intervals are fairly wide. Even using the upper confidence limit for sensitivity, however, the worst-performing breathalyser would still have sensitivity under 50 per cent. Our sample of participants obtained from colleges and pubs may not be representative of the population who purchase personal breathalysers, particularly in age group and quantity of alcohol consumed and socio-demographics; although this might affect sensitivity of self-reported alcohol consumption, it should not impact significantly on the overall accuracy of the breathalyser devices. We tested the index devices on the night of drinking and the generalizability to "morning-after" use may be unclear; blood and breath alcohol concentrations decline with time after drinking alcohol at similar rates and maintain high correlation,[11] but it is possible that diagnostic accuracy of personal breathalysers may differ at lower alcohol concentrations.

Use of blood alcohol as a gold standard was not possible in this pragmatic, field study. We chose as our reference standard the Dräger Alcotest 6510 which has passed a home office testing protocol requiring error less than 10% in all readings [12,13] and is also an evidential device for criminal prosecutions in the US. The devices currently approved for evidential use in the UK are not portable and were therefore not suitable for our study setting. For convenience we selected index devices easily available in the UK for testing, however the same devices calibrated for different legal limits are sold outside of the UK, and other devices sold elsewhere use similar technology and are similarly priced. Therefore, while we cannot directly apply our results to other countries, we would anticipate similar findings elsewhere.

We have not found other studies testing the accuracy of personal breathalysers. Studies in the US of college student parties have found a mean BAC of 0.077% (standard deviation 0.063%) (for context, the UK driving limit is 0.08%).[14] A Canadian study found a linear relationship between self-reported alcohol consumption and BAC in Emergency Room attendees, up to 7 drinks.[15] However, a US study of college students which compared estimated BAC to measured BAC, found that students tended to over-estimate their levels of consumption when surveyed in the midst of a night of drinking.[16] We did not attempt to convert self-reported alcohol consumption to BAC and only recorded total quantity consumed, which would have had differential effects between individuals dependent on weight, sex and the time over which the drinks were consumed. However, despite the variation these factors would introduce into BAC, self-reported consumption up to 5 units

was still a more sensitive test than the breathalysers tested, and BAC is known to correlate poorly with symptoms of intoxication.[17]

Our research suggests that some personal breathalysers available for sale to the public are not fit for purpose. The use of inaccurate information from breathalysers thought to be accurate could have catastrophic safety implications for drivers. The fact that these devices are sold in well-established pharmacies including national chains does not guarantee sufficient accuracy for safe use. Medical and measurement devices may carry regulatory approvals such as CE- or NF-marking, but this does not appear to correlate with accuracy, and this raises wider questions over how this marking may be perceived by users. A derivative device of the worst-performing breathalyser in our study is widely sold for use in France as part of the new law requiring breathalysers to be carried when driving, and has French NF-approval.[4] Although results from our study cannot be directly applied to the lower French driving limit of 0.05g/dl and a derivative device, it questions the utility of the new law which on the one hand may improve public awareness of drink-driving in general, but also risks ill-informed driving decisions based on inaccurate results from a personal breathalyser. Replication of our results in other settings and with other breathalysers could further inform policy makers planning to introduce similar laws in other jurisdictions, and explore the characteristics of the population who purchase personal breathalysers and how they use the results obtained. Finally, our research raises worrying questions about the level of scrutiny that medical tests intended for sale to the general public undergo in Europe, and raises wider concerns about how diagnostic accuracy in particular is evaluated, and whether any further field evaluations are required for intended users, perceptions of accuracy of such devices and how use of such devices interacts with medical testing in other health care settings.

Competing interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

The authors would like to thank the study participants, and those colleges of the University of Oxford and public houses in Oxford which permitted them to recruit on their premises and Michael Ashdown for assistance with legal literature.

Contributors

All authors contributed to study design, data collection, data interpretation and helped write the paper. RJS conceived the idea for the study. HFA, EAS and RJS analysed the data. HFA is guarantor. All authors approved the final version of the manuscript.

Ethical approval

The study was approved by the University of Oxford Medical Sciences Division Interdivisional Research Ethics Committee (reference MSD/IDREC/2011/13). All participants gave informed consent before taking part. No remuneration or incentives were provided to participants for taking part.

Funding

This paper summarises independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0407-10347). The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Sponsorship

The University of Oxford acted as study sponsor and had no role in study design, data collection, analysis or interpretation, writing of the paper or the decision to submit for publication. All authors were independent from funders and sponsors.

Data sharing

All authors had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. Full data are available from the corresponding author on request. Consent for data sharing was not obtained but the presented data are anonymised and risk of identification is low.

Transparency declaration

The lead author and the manuscript's guarantor affirm 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 any discrepancies from the study as planned have been explained.

FIGURE LEGENDS

Figure 1: Study flow chart. All participants (except exclusions detailed in figure) were tested with the Dräger Alcotest 6510 (R), the Alcosense Elite (IM), and one of either the Dräger Alco-Check (ISD) or the Alcosense Single (ISA). The order of undertaking each breathalyser, and the selection of ISD or ISA was determined by randomisation. *3 participants left before analysis therefore results for participant estimation of ISA are only available for 97 participants.

Figure 2: ROC curve of self-reported alcohol consumption in UK units, with comparative sensitivity and specificity for breathalysers tested.

REFERENCES

1. Peden M, Scurfield R, Sleet D, et al. World report on road traffic injury prevention. Geneva: World Health Organization 2004.
2. Global Road Safety Partnership. Drinking and driving: a road safety manual for decision-makers and practitioners. Geneva: World Health Organization 2007.
3. Légifrance: Le Service Public pour la Diffusion du Droit: Décret n° 2012-284 du 28 février 2012 relatif à la possession obligatoire d'un éthylotest par le conducteur d'un véhicule terrestre à moteur http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000025417826&categorieLien=id [accessed May 2014].
4. Halfords store website. Alcosense Single breathalyser http://www.halfords.com/webapp/wcs/stores/servlet/product_storeId_10001_catalogId_10151_productId_871069_langId_-1_categoryId_226211 [accessed May 2014]
5. UK Breathalysers website. Draeger Police 6510. http://www.ukbreathalysers.com/2_alcohol/test_others/prob/6510.html [accessed May 2014].
6. Gibb KA, Yee AS, Johnston CC, et al. Accuracy and usefulness of a breath alcohol analyzer. *Ann Emerg Med* 1984;13:516-20.
7. Road Traffic Act 1988 Chapter 52 Part I Sections 4-11. Motor vehicles: drink and drugs. Ministry of Justice, UK http://www.legislation.gov.uk/ukpga/1988/52/part/I/crossheading/motor-vehicles-drink-and-drugs [accessed May 2014].
8. Dräger Alcotest 6510 product specification 2012.
9. McCance RA, Widdowson, EM. McCance and Widdowson's The Composition of Foods. 6th edition compiled by the Food Standards Agency and Institute of Food Research, Cambridge, UK: Royal Society of Chemisty 2002.
10. Newcombe RG, Altman, DG. Proportions and their differences. In: Altman DG, Bryant TN, Gardner MJ, eds. Statistics With Confidence. 2nd edition ed. London: BMJ Books 2000:46-54.
11. Jones AW, Andersson L. Comparison of ethanol concentrations in venous blood and end-expired breath during a controlled drinking study. *Forensic Sci Int* 2003;132:18-25.

12. Parliamentary Under-Secretary of State, Home Office. The Preliminary Breath Test Device Approval 2005. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/117005/draeger-lion-envitec-alcoquant.pdf [accessed August 2014].
13. Home Office and Forensic Science Service. Breath Alcohol Screening Devices: A guide to type-approval procedures for breath alcohol screening devices used for law enforcement in Great Britain (Version 1.0, 30/06/2004). https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/117043/breathalcoholscreening.pdf [accessed August 2014].
14. Clapp JD, Min JW, Shillington AM, et al. Person and environment predictors of blood alcohol concentrations: a multi-level study of college parties. *Alcohol Clin Exp Res* 2008;32:100-7.
15. Bond J, Ye Y, Cherpitel CJ, et al. The relationship between self-reported drinking and BAC level in emergency room injury cases: is it a straight line? *Alcohol Clin Exp Res* 2010;34:1118-25.
16. Kraus CL, Salazar NC, Mitchell JR, et al. Inconsistencies between actual and estimated blood alcohol concentrations in a field study of college students: do students really know how much they drink? *Alcohol Clin Exp Res* 2005;29:1672-6.
17. Olson KN, Smith SW, Kloss JS, et al. Relationship between blood alcohol concentration and observable symptoms of intoxication in patients presenting to an emergency department. *Alcohol Alcohol* 2013;48:386-9.

Breathalyser	Total participants	Correctly positive by reference standard	Correctly negative by reference standard	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)
Alcosense Elite multi-use	205	34/38	107/167	89.5 (75.9 to 95.8)	64.1 (56.6 to 71.0)	36.2 (27.2 to 46.3)	96.4 (91.1 to 98.6)
Dräger Alco-Check single-use	108	18/19	45/89	94.7 (75.4 to 99.1)	50.6 (40.4 to 60.7)	29.0 (19.2 to 41.3)	97.8 (88.7 to 99.6)
Alcosense Single single-use	100	5/19	79/81	26.3 (11.8 to 48.8)	97.5 (91.4 to 99.3)	71.4 (35.9 to 91.8)	85.0 (76.3 to 90.8)

Table 1: Diagnostic accuracy of index breathalysers compared to reference police breathalyser, using researcher interpretation of single-use breathalysers. 95% confidence intervals shown in parentheses.

Threshold of self-reported alcohol consumption (UK units)	Total participants	Correctly positive by reference standard	Correctly negative by reference standard	Sensitivity (%)	Specificity (%)
1	207	38/38	9/169	100.0 (90.8 to 100.0)	5.3 (2.8 to 9.8)
2	207	38/38	23/169	100.0 (90.8 to 100.0)	13.6 (9.2 to 19.6)
3	207	38/38	52/169	100.0 (90.8 to 100.0)	30.8 (24.3 to 38.1)
4	207	38/38	71/169	100.0 (90.8 to 100.0)	42.0 (34.8 to 49.5)
5	207	37/38	84/169	97.4 (86.5 to 99.5)	49.7 (42.3 to 57.2)
6	207	35/38	105/169	92.1 (79.2 to 97.3)	62.1 (54.6 to 69.1)
7	207	34/38	117/169	89.5 (75.9 to 95.8)	69.2 (61.9 to 75.7)
8	207	29/38	130/169	76.3 (60.8 to 87.0)	76.9 (70.0 to 82.6)
9	207	26/38	142/169	68.4 (52.5 to 80.9)	84.0 (77.7 to 88.8)
10	207	20/38	148/169	52.6 (37.3 to 67.5)	87.6 (81.8 to 91.7)

Table 2: Diagnostic accuracy of self-reported alcohol consumption compared to reference police breathalyser. 1 UK unit is equivalent to 10ml or 8g alcohol (N.B. total 207 partcipants due to missing data on alcohol consumption for one participant). 95% confidence intervals shown in parentheses.

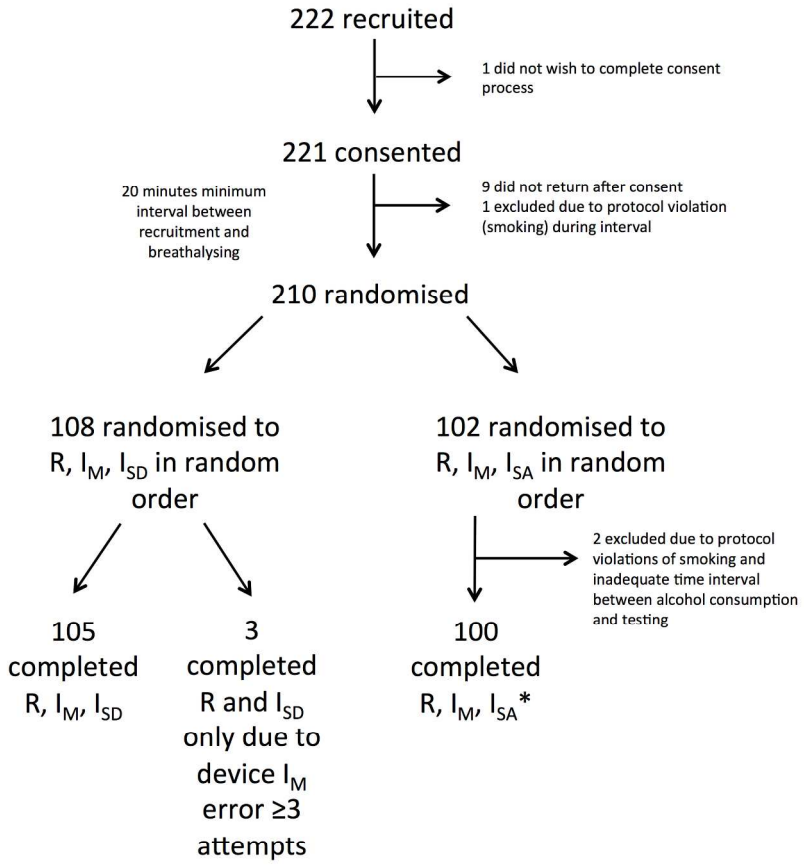
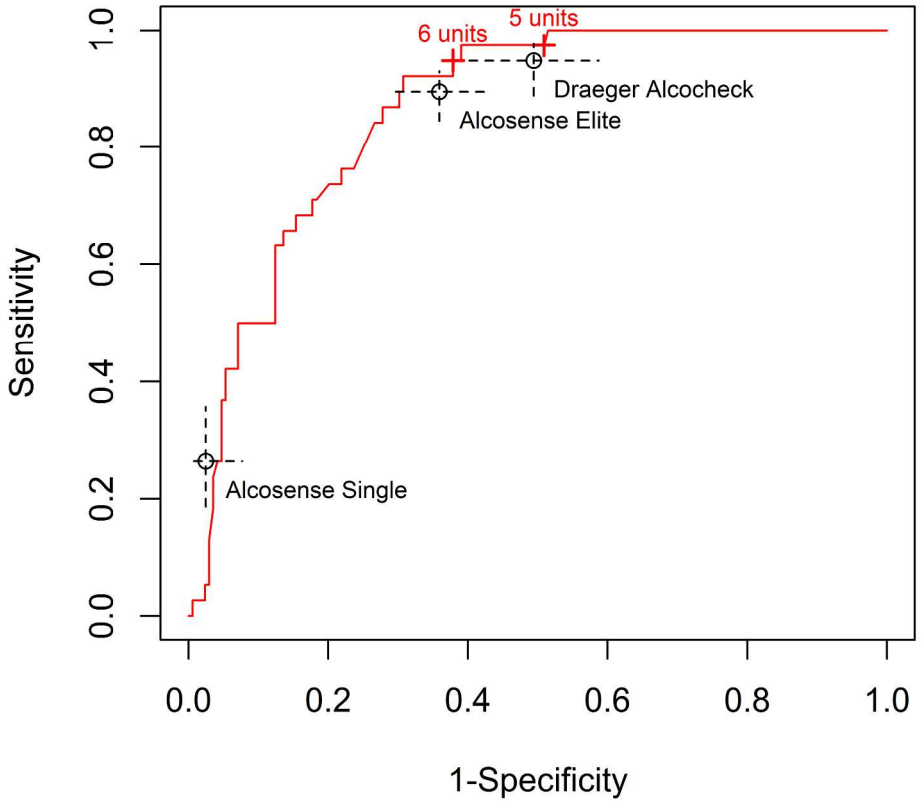


Figure 1: Study flow chart. All participants (except exclusions detailed in figure) were tested with the Dräger Alcotest 6510 (R), the Alcosense Elite (IM), and one of either the Dräger Alco-Check (ISD) or the Alcosense Single (ISA). The order of undertaking each breathalyser, and the selection of ISD or ISA was determined by randomisation. *3 participants left before analysis therefore results for participant estimation of ISA are only available for 97 participants.
190x254mm (300 x 300 DPI)



ROC curve of self-reported alcohol consumption in UK units, with comparative sensitivity and specificity for breathalysers tested.
127x127mm (600 x 600 DPI)

Web appendix: full 2x2 table data and diagnostic accuracy results for researcher and participant estimation and sensitivity analyses

Main analysis

Alcosense Elite

		Reference		
		Positive	Negative	Total
Index	Positive	34	60	94
	Negative	4	107	111
	Total	38	167	205

		Lower limit CI	*Upper limit CI*
Sensitivity	89.47%	*75.87%*	*95.83%*
Specificity	64.07%	*56.55%*	*70.96%*
PPV	36.17%	*27.18%*	*46.25%*
NPV	96.40%	*91.10%*	*98.59%*

Dräger Alco-Check

		Reference		
		Positive	Negative	Total
Index	Positive	18	44	62
	Negative	1	45	46
	Total	19	89	108

		Lower limit CI	*Upper limit CI*
Sensitivity	94.74%	*75.36%*	*99.06%*
Specificity	50.56%	*40.37%*	*60.71%*
PPV	29.03%	*19.22%*	*41.29%*
NPV	97.83%	*88.66%*	*99.62%*

Alcosense Single

		Reference		
		Positive	Negative	Total
Index	Positive	5	2	7
	Negative	14	79	93
	Total	19	81	100

		Lower limit CI	*Upper limit CI*
Sensitivity	26.32%	*11.81%*	*48.79%*
Specificity	97.53%	*91.44%*	*99.32%*
PPV	71.43%	*35.89%*	*91.78%*
NPV	84.95%	*76.30%*	*90.82%*

Participant estimation

Alcosense Elite

Not applicable as no participant estimation
(i.e. same as main analysis)

Dräger Alco-Check

		Reference		
		Positive	Negative	Total
Index	Positive	18	42	60
	Negative	1	47	48
	Total	19	89	108

		Lower limit CI	*Upper limit CI*
Sensitivity	94.74%	*75.36%*	*99.06%*
Specificity	52.81%	*42.54%*	*62.85%*
PPV	30.00%	*19.90%*	*42.51%*
NPV	97.92%	*89.10%*	*99.63%*

Alcosense Single

		Reference		
		Positive	Negative	Total
Index	Positive	3	4	7
	Negative	15	75	90
	Total	18	79	97

		Lower limit CI	*Upper limit CI*
Sensitivity	16.67%	*11.81%*	*48.79%*
Specificity	94.94%	*91.44%*	*99.32%*
PPV	42.86%	*35.89%*	*91.78%*
NPV	83.33%	*76.30%*	*90.82%*

Participant estimation - sensitivity analysis excluding colour blind person (A/168)

Alcosense Elite

Not applicable as no participant estimation
(i.e. same as main analysis)

Dräger Alco-Check

		Reference		
		Positive	Negative	Total
Index	Positive	18	41	59
	Negative	1	47	48
	Total	19	88	107

		Lower limit CI	*Upper limit CI*
Sensitivity	94.74%	*75.36%*	*99.06%*
Specificity	53.41%	*43.06%*	*63.47%*
PPV	30.51%	*20.25%*	*43.15%*
NPV	97.92%	*89.10%*	*99.63%*

Alcosense Single

Not applicable as colour blind person not in this group
(i.e. same as participant estimation above)

Researcher estimation - sensitivity analysis excluding suspect protocol violation (A/111)

Alcosense Elite

		Reference		
		Positive	Negative	Total
Index	Positive	34	60	94
	Negative	3	107	110
	Total	37	167	204

		Lower limit CI	*Upper limit CI*
Sensitivity	91.89%	*78.70%*	*97.20%*
Specificity	64.07%	*56.55%*	*70.96%*
PPV	36.17%	*27.18%*	*46.25%*
NPV	97.27%	*92.29%*	*99.07%*

Dräger Alco-Check

Not applicable as protocol violation person not in this group
(i.e. same as main analysis)

Alcosense Single

		Reference		
		Positive	Negative	Total
Index	Positive	5	2	7
	Negative	13	79	92
	Total	18	81	99

		Lower limit CI	*Upper limit CI*
Sensitivity	27.78%	*12.50%*	*50.87%*
Specificity	97.53%	*91.44%*	*99.32%*
PPV	71.43%	*35.89%*	*91.78%*
NPV	85.87%	*77.31%*	*91.55%*

Researcher estimation - sensitivity analysis excluding potential faults with devices (A/165 and A/179)

Alcosense Elite

Not applicable as no potential errors with Alcosense Elite
(i.e. same as main analysis)

Dräger Alco-Check

		Reference		
		Positive	Negative	Total
Index	Positive	18	44	62
	Negative	0	45	45
	Total	18	89	107

		Lower limit CI	*Upper limit CI*
Sensitivity	100.00%	*82.41%*	*100.00%*
Specificity	50.56%	*40.37%*	*60.71%*
PPV	29.03%	*19.22%*	*41.29%*
NPV	100.00%	*92.13%*	*100.00%*

Alcosense Single

		Reference		
		Positive	Negative	Total
Index	Positive	5	2	7
	Negative	14	78	92
	Total	19	80	99

		Lower limit CI	*Upper limit CI*
Sensitivity	26.32%	*11.81%*	*48.79%*
Specificity	97.50%	*91.34%*	*99.31%*
PPV	71.43%	*35.89%*	*91.78%*
NPV	84.78%	*76.06%*	*90.71%*

Diagnostic accuracy Study of three alcohol breathalyzers marketed for sale to the public

Aghdam HF et al

Table 1. STARD checklist for the reporting of studies of diagnostic accuracy.

Section and Topic	Item #		On page #
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	1
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	4
METHODS		Describe	
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where the data were collected.	5
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	5
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4? If not, specify how participants were further selected.	5
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	5
Test methods	7	The reference standard and its rationale.	
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	5
	9	Definition of and rationale for the units, cutoffs and/or categories of the results of the index tests and the reference standard.	6
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard.	5
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	5/6
Statistical methods	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	6
	13	Methods for calculating test reproducibility, if done.	N/a
RESULTS		Report	
Participants	14	When study was done, including beginning and ending dates of recruitment.	5
	15	Clinical and demographic characteristics of the study population (e.g. age, sex, spectrum of presenting symptoms, comorbidity, current treatments, recruitment centers).	7
	16	The number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test (a flow diagram is strongly recommended).	7/ Figure 1
Test results	17	Time interval from the index tests to the reference standard, and any treatment administered between.	6
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	7
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	7/ Table 1/ Web appendix
	20	Any adverse events from performing the index tests or the reference standard.	7
Estimates	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	7/Table 1/ Web appendix
	22	How indeterminate results, missing responses and outliers of the index tests were handled.	6/7/ Web appendix
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	7/ Web appendix
	24	Estimates of test reproducibility, if done.	N/a
DISCUSSION	25	Discuss the clinical applicability of the study findings.	8-10

BMJ Open

Diagnostic accuracy study of three alcohol breathalysers marketed for sale to the public

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-005811.R2
Article Type:	Research
Date Submitted by the Author:	19-Nov-2014
Complete List of Authors:	Ashdown, Helen; University of Oxford, Nuffield Department of Primary Care Health Sciences Fleming, Susannah; University of Oxford, Nuffield Department of Primary Care Health Sciences Spencer, Elizabeth; University of Oxford, Primary Care Health Sciences Thompson, Matthew; University of Oxford, Primary Care Health Sciences Stevens, Richard; University of Oxford, Department of Primary Health Care
Primary Subject Heading:	Diagnostics
Secondary Subject Heading:	Health policy, Addiction
Keywords:	BIOTECHNOLOGY & BIOINFORMATICS, PUBLIC HEALTH, TOXICOLOGY, PREVENTIVE MEDICINE

Diagnostic accuracy study of three alcohol breathalysers marketed for sale to the public

Helen F Ashdown[1], Susannah Fleming[1], Elizabeth A Spencer[1], Matthew J Thompson[1], Richard J Stevens[1]

[1] Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, United Kingdom

Helen F Ashdown
Clinical Researcher in General Practice

Susannah Fleming
Research Fellow

Elizabeth A Spencer
Nutritional Epidemiologist

Matthew J Thompson
Researcher

Richard J Stevens
University Research Lecturer in Medical Statistics

Correspondence to:
HF Ashdown
helen.ashdown@phc.ox.ac.uk

Word count (main text): 2745

ABSTRACT

Objectives

To assess the diagnostic accuracy of 3 personal breathalyser devices available for sale to the public marketed to test safety to drive after drinking alcohol

Design

Prospective comparative diagnostic accuracy study comparing two single-use breathalysers and one digital multi-use breathalyser (index tests) to a police breathalyser (reference test).

Setting

Establishments licensed to serve alcohol in a UK city

Participants

Of 222 participants recruited, 208 were included in the main analysis. Participants were eligible if they were 18 years old or over, had consumed alcohol and were not intending to drive within the following six hours.

Outcome measures

Sensitivity and specificity of the breathalysers for the detection of being at or over the UK legal driving limit (35 micrograms/100ml breath alcohol concentration).

Results

18% of participants (38/208) were at or over the UK driving limit according to the police breathalyser. The digital multi-use breathalyser had a sensitivity of 89.5% (95% CI 75.9 to 95.8%) and a specificity of 64.1% (95% CI 56.6 to 71.0%). The single-use breathalysers had a sensitivity of 94.7% (95% CI 75.4 to 99.1%) and 26.3% (95% CI 11.8 to 48.8%), and a specificity of 50.6% (95% CI 40.4 to 60.7%) and 97.5% (95% CI 91.4 to 99.3%) respectively. Self-reported alcohol consumption threshold of 5 UK units or fewer had a higher sensitivity than all personal breathalysers.

Conclusions

Sensitivities of alcohol breathalysers marketed to motorists in the UK for driving decisions varied between 26% and 95%, corresponding to approximately three people in four, and one person in twenty, who are falsely reassured when over the limit. Our study suggests that some devices are not sufficiently sensitive to test safety to drive after drinking alcohol, and that policy-makers should reconsider the regulatory frameworks that allow the manufacture and widespread marketing of a medical device with low sensitivity for a decision with potentially catastrophic consequences.

Strengths and limitations of this study

- Personal breathalysers are available for sale to the public in pharmacies for assessing safety to drive following consuming alcohol, some very inexpensively, and in some jurisdictions is now promoted by law.
- Accuracy of personal breathalyser devices has never previously been studied.
- This study tested diagnostic accuracy of these devices in a real-life setting including participant estimation of readings.
- Limitations include the uncontrolled environment of public houses and bars, use of a pragmatic reference standard, and wide confidence intervals for some results due to low prevalence of those over the driving limit.
- However, our conclusions for the worst performing device are robust even against this uncertainty, since even the upper confidence limit for sensitivity of the worst performing device (48.8%) would still mean that about one in two individuals would be falsely reassured and this could lead to potentially dangerous driving decisions.

INTRODUCTION

Road traffic collisions (RTCs) are the second leading cause of death worldwide amongst people aged 5-29 years, estimated at 1.2 million deaths each year, and forecast to rise further by 2020.[1] Consumption of alcohol is an important factor influencing the likelihood and severity of RTCs, and has been found to be a causal factor in 17-40% of RTCs worldwide.[2] The risk of an RTC increases rapidly with increasing blood alcohol concentration, with relative risk rising significantly beyond a blood alcohol concentration (BAC) of 0.04g alcohol per 100ml blood (g/dl) to reach a relative risk of approximately 5 at 0.10g/dl.[2] Introduction of maximum legal BAC limits for driving, which vary from 0.02g/dl (e.g. Russia) to 0.15g/dl (e.g. Uganda), has been effective in reducing alcohol-related injuries and deaths.[1]

Alcohol breathalysers, which have long been available to law enforcement agencies, are now marketed direct to consumers, for example in some UK pharmacies and motoring stores, to test safety to drive following drinking alcohol, including the morning after. In July 2012, it became a legal requirement for drivers in France to carry a personal breathalyser at all times.[3] Devices marketed to consumers are order of magnitudes lower in cost than those intended for law enforcement: for example at the time of writing the Alcosense Single is more than 300 times cheaper than the Dräger 6510 Home Office certified police breathalyser.[4, 5]

The theoretical accuracy of breath alcohol measurement as a surrogate for blood alcohol measurement has been well-studied,[6] and in the UK breath alcohol forms part of the prescribed legal limit, along with blood and urine alcohol concentration.[7] However to our knowledge the accuracy of devices currently marketed to the consumer/motorist has not been studied. Many such devices carry regulatory approvals such as the Conformité Européenne (CE), French NF certification and British Standard (BSI Kitemark) but in general such marks are statements of engineering quality rather than diagnostic accuracy. We therefore aimed to assess the diagnostic accuracy of personal breathalysers compared to a police breathalyser for detection of being at or over the UK legal driving limit.

METHODS

Index tests and reference standard

We selected for study as index tests the Alcosense Single, the UK counterpart of an NF-approved device widely sold for motorists in France; the Dräger Alco-Check, as a comparable single-use device from a competing manufacturer; and the Alcosense Elite, as an example of a digital multi-use device readily available from pharmacies, high street and online stores (Boots, Halfords, Amazon and others). We selected as reference standard the Dräger Alcotest 6510 device. This has Home Office approval and is standard issue to UK police for use at the roadside,[5] and is approved in the US as an evidential breath testing device (National Highway Traffic Safety Administration Standard 49 FR 48854) meaning readings can be used as evidence for prosecution of drink driving offences. Manufacturer information states measurement precision as ±0.008mg/L or 1.7% of measurement value.[8] The Dräger Alcotest 6510 and Alcosense Elite devices come with instruction manuals including details of breathalyser operating technique and result interpretation. The single-use devices contain information on their packaging and an enclosed sheet of paper including details of breathalyser operating technique and result interpretation. These manuals and leaflets recommended at least 15 minutes (Dräger Alcocheck), at least 20 minutes (Dräger 6510 and Alcosense Single), or at least 30 minutes (Alcosense Elite) elapse between alcohol consumption and use. All personal breathalyser instructions clearly state that any amount of alcohol even below the limit can impair driving ability.

Study participants

We recruited participants from establishments licensed to serve alcohol in the city centre of Oxford, UK, including college bars and public houses. In the absence of prior data with which to estimate sample size, we decided to recruit 200 participants, which would allow us to report the prevalence of intoxication with small standard error (maximum 3.5%). Recruitment took place on evenings during the period from October 2012 to January 2013: 11 study evenings were required to reach a sample size greater than 200. Participants were eligible if they were 18 years or over, had consumed alcohol, and were not intending to drive within the following six hours.

Recruitment and consent

Individuals in these establishments were informally approached by a member of the research team and given preliminary information about the study. Potentially eligible interested participants were then asked to sit with the research team, at tables reserved within the same premises, where they received a full description of the study, were asked to read the study literature, and given an opportunity to ask further questions. Eligibility was checked and written consent taken. Participants were given a card with contact details to use in the event of withdrawing consent subsequently (for example, the next day): in the event, no participants used this option to withdraw consent.

Study procedures

The research team consisted of between two and four individuals, who followed the reference breathalyser manufacturers' written instructions in directing their use: a minimum of 20 minutes was enforced between recruitment and using the breathalyser devices, and participants were asked not to drink further alcohol, smoke, use mouthwash, or drink fruit juice during this period. Participants were provided with water and asked to take at least one sip in order to clear any residual alcohol from the upper airway. During this 20 minute period basic demographic details (age and sex) and reported alcoholic drinks consumption during the preceding 12 hours were recorded by the researchers. Participants were not required to remain under observation during the 20 minute period.

Participants then used three breathalysers, at intervals of at least one minute, at the study tables under the supervision of a research team member. Each participant used, in random order, the reference standard, the Alcosense Elite multi-use and a randomly selected single use breathalyser. Randomisation was carried out in advance by study number in random permuted blocks.

A member of the research team recorded digital outputs of the multi-use and reference breathalysers. The single-use breathalysers require subjective assessment of colour of crystals after use: we recorded assessments by both researcher (primary outcome) and participant (secondary outcome). Participants were blinded to the researcher assessment and to the reference result.

Statistical analysis

We calculated sensitivity, specificity, positive and negative predictive values of the breathalysers for the outcome of being at or over the current UK legal driving limit[7] of 35 micrograms/100ml or 0.8‰BAC according to the reference standard. Self-reported alcohol consumption, converted to UK alcohol units (1 unit = 8g alcohol) using nutritional tables,[9] was also assessed for diagnostic accuracy of being at or over the UK legal driving limit, and an ROC analysis undertaken to assess different unit thresholds. Statistical calculations were carried out using standard methods[10] and the ROC analysis using Stata (Release 11). Minor protocol violations were discussed amongst the research team, with sensitivity analyses performed to determine whether there was any difference in overall results due to inclusion or exclusion of these. Participants with missing data (for example sex, self-reported alcohol consumption or participant estimation of result) were included except for analyses involving the missing component itself.

RESULTS

A total of 208 participants were included in the main analysis (see Figure 1 flowchart), of whom 148/207 (71.5%) were male and with a median age of 20. Participants reported having consumed a median of 6 UK units of alcohol (range 1 to 25), equivalent to a median of 46g (range 8 to 204) alcohol.

38 participants (18.3%, 95% CI 11.7 to 27.4%) were at or above the current UK driving limit of 35 μg/100ml according to the reference breathalyser. Table 1 compares performance of the three index breathalysers at detecting those at or over the UK limit. Compared to the reference breathalyser, the Alcosense Elite multi-use breathalyser had a sensitivity of 89.5% (95% CI 75.9 to 95.8%), the Dräger Alco-Check single-use breathalyser had a sensitivity of 94.7% (95% CI 75.4 to 99.1%) and the Alcosense Single breathalyser had a sensitivity of 26.3% (95% CI 11.8 to 48.8%). When analyses were repeated using the participant's interpretation of colour change in the single-use breathalysers instead of researcher's interpretation, sensitivity was 94.7% (95% CI 75.4 to 99.1%) for the Dräger Alco-Check (i.e. identical to researcher estimation) and 16.7% (95% CI 11.8 to 48.8%) for the Alcosense Single (see Appendix 1 supplementary material for full data for participant estimation).

We conducted three sensitivity analyses in turn (i) excluding two results where the participants were suspected to have incorrectly used the device, (ii) a colour-blind participant who may have had difficulty interpreting the colour change of crystals, and (iii) a participant who was suspected to have violated the protocol (consumed alcohol between use of the three devices) showed minimal difference to results overall (see Appendix 1) although the sensitivity of the Dräger Alco-Check increased to 100% in the latter sensitivity analysis. There were no reported adverse events from using any of the breathalyser devices.

Because there is no single standard threshold for driving decisions based on self-estimated alcohol consumption, we calculated sensitivity and specificity for all alcohol unit thresholds (Table 2) and plotted them in ROC space (Figure 2).

DISCUSSION

We have shown that breathalysers available for sale to the public for personal use vary considerably in their performance in detecting being at or over a legal driving limit. Two of the devices tested, the Alcosense Elite digital multi-use breathalyser and Dräger Alco-Check single-use breathalyser had a sensitivity of approximately 90 and 95 per cent respectively. However, even this means that approximately 1 in 20 people over the driving limit would be falsely reassured by these tests. The third device, the Alcosense Single single-use breathalyser, had a sensitivity of only 26 per cent, meaning that only approximately one in four individuals over the legal limit would be identified by this device. Participants (rather than researchers) interpreting results, which is what would occur in real life, reduced sensitivity further to only 17 per cent. This device has a correspondingly high specificity, but specificity is not the safety-critical aspect of performance of a device assessing safety to drive. Surprisingly, we found that self-reported consumption of alcohol was a more sensitive test for being at or over the legal driving limit than the three breathalysers tested for up to 5 UK units of alcohol consumed. None of the devices outperformed simple recall of amount of alcohol consumed up to 5 UK units of alcohol.

Strengths of our study include testing participants in a "real-world" environment including assessment of participants' estimation of breathalyser readings, which enables generalizability to everyday life. However, this pragmatic approach also brings limitations in that the setting in college bars and public houses could not be rigorously controlled and it is possible that the environment, for example poor lighting or unknown protocol violation through drinking alcohol or smoking immediately prior to testing, could have introduced some inaccuracies. Operating the three breathalysers in close succession and randomisation of order of use of the breathalysers should have reduced the impact of this. As discussed above, we may have underestimated the sensitivity of the Dräger Alco-Check because of a suspected protocol violation. Three participants inadvertently took the breathalysers in a different order to that planned: we decided to include these participants because the aim was to have an overall variety in breathalyser order and this was unlikely to have any impact for such a small number of participants. The use of a 20 minute rather than 30 minute minimum time after drinking alcohol was not strictly adherent to the manufacturer's instructions for the Alcosense Elite, and so it is possible that this may have affected results for this breathalyser.

Participants were blinded to their result on the reference breathalyser until after completing estimation of colour change, however researchers were not and it is possible that this could have introduced some bias. We met our planned recruitment target, but the prevalence of those over the legal limit by the reference standard means that our confidence intervals are fairly wide. Even using the upper confidence limit for sensitivity, however, the worst-performing breathalyser would still have sensitivity under 50 per cent. Our sample of participants obtained from colleges and pubs may not be representative of the population who purchase personal breathalysers, particularly in age group and quantity of alcohol consumed and socio-demographics; although this might affect sensitivity of self-reported alcohol consumption, it should not impact significantly on the overall accuracy of the breathalyser devices. We tested the index devices on the night of drinking and in a bar environment and the generalizability to other uses such as "morning-after" use may be unclear; blood and breath alcohol concentrations decline with time after drinking alcohol at similar rates and maintain high correlation,[11] but it is possible that diagnostic accuracy of personal breathalysers may differ at lower alcohol concentrations.

Use of blood alcohol as a gold standard was not possible in this pragmatic, field study. We chose as our reference standard the Dräger Alcotest 6510 which has passed a home office testing protocol requiring error less than 10% in all readings [12,13] and is also an evidential device for criminal prosecutions in the US. The devices currently approved for evidential use in the UK are not portable and were therefore not suitable for our study setting. For convenience we selected index devices easily available in the UK for testing, however the same devices calibrated for different legal limits are sold outside of the UK, and other devices sold elsewhere use similar technology and are similarly priced. Therefore, while we cannot directly apply our results to other countries, we would anticipate similar findings elsewhere.

We have not found other studies testing the accuracy of personal breathalysers. Studies in the US of college student parties have found a mean BAC of 0.077% (standard deviation 0.063%) (for context, the UK driving limit is 0.08%).[14] A Canadian study found a linear relationship between self-reported alcohol consumption and BAC in Emergency Room attendees, up to 7 drinks.[15] However, a US study of college students which compared estimated BAC to measured BAC, found that students tended to over-estimate their levels of consumption when surveyed in the midst of a night of drinking.[16] We did not attempt to convert self-reported alcohol consumption to BAC and only recorded total quantity

consumed, which would have had differential effects between individuals dependent on weight, sex and the time over which the drinks were consumed. However, despite the variation these factors would introduce into BAC, self-reported consumption up to 5 units was still a more sensitive test than the breathalysers tested, and BAC is known to correlate poorly with symptoms of intoxication.[17]

Our research suggests that some personal breathalysers available for sale to the public are not sufficiently sensitive to test safety to drive after drinking alcohol, where use of inaccurate information from breathalysers thought to be accurate could have catastrophic safety implications for drivers. The fact that these devices are sold in well-established pharmacies including national chains does not guarantee sufficient accuracy for safe use. Medical and measurement devices may carry regulatory approvals such as CE- or NF-marking, but this does not appear to correlate with accuracy, and this raises wider questions over how this marking may be perceived by users. A derivative device of the worst-performing breathalyser in our study is widely sold for use in France as part of the new law requiring breathalysers to be carried when driving, and has French NF-approval.[4] Although results from our study cannot be directly applied to the lower French driving limit of 0.05g/dl and a derivative device, they question the utility of the new law which on the one hand may improve public awareness of drink-driving in general, but also risks ill-informed driving decisions based on inaccurate results from a personal breathalyser. Replication of our results in other settings and with other breathalysers could further inform policy makers planning to introduce similar laws in other jurisdictions, and explore the characteristics of the population who purchase personal breathalysers and how they use the results obtained. Finally, our research raises worrying questions about the level of scrutiny that medical tests intended for sale to the general public undergo in Europe, and raises wider concerns about how diagnostic accuracy in particular is evaluated, and whether any further field evaluations are required for intended users, perceptions of accuracy of such devices and how use of such devices interacts with medical testing in other health care settings.

Competing interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

The authors would like to thank the study participants, and those colleges of the University of Oxford and public houses in Oxford which permitted them to recruit on their premises and Michael Ashdown for assistance with legal literature.

Contributors

All authors contributed to study design, data collection, data interpretation and helped write the paper. RJS conceived the idea for the study. HFA, EAS and RJS analysed the data. HFA is guarantor. All authors approved the final version of the manuscript.

Ethical approval

The study was approved by the University of Oxford Medical Sciences Division Interdivisional Research Ethics Committee (reference MSD/IDREC/2011/13). All participants gave informed consent before taking part. No remuneration or incentives were provided to participants for taking part.

Funding

This paper summarises independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0407-10347). The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Sponsorship

The University of Oxford acted as study sponsor and had no role in study design, data collection, analysis or interpretation, writing of the paper or the decision to submit for publication. All authors were independent from funders and sponsors.

Data sharing

All authors had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. Full data are available from the corresponding author on request. Consent for data sharing was not obtained but the presented data are anonymised and risk of identification is low.

Transparency declaration

The lead author and the manuscript's guarantor affirm 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 any discrepancies from the study as planned have been explained.

Figure Legends

Figure 1: Study flow chart. All participants (except exclusions detailed in figure) were tested with the Dräger Alcotest 6510 (R), the Alcosense Elite (IM), and one of either the Dräger Alco-Check (ISD) or the Alcosense Single (ISA). The order of undertaking each breathalyser, and the selection of ISD or ISA was determined by randomisation. *3 participants left before analysis therefore results for participant estimation of ISA are only available for 97 participants.

Figure 2: ROC curve of self-reported alcohol consumption in UK units, with comparative sensitivity and specificity for breathalysers tested.

REFERENCES

1. Peden M, Scurfield R, Sleet D, et al. World report on road traffic injury prevention. Geneva: World Health Organization 2004.
2. Global Road Safety Partnership. Drinking and driving: a road safety manual for decision-makers and practitioners. Geneva: World Health Organization 2007.
3. Légifrance: Le Service Public pour la Diffusion du Droit: Décret n° 2012-284 du 28 février 2012 relatif à la possession obligatoire d'un éthylotest par le conducteur d'un véhicule terrestre à moteur http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000025417826&categorieLien=id [accessed May 2014].
4. Halfords store website. Alcosense Single breathalyser http://www.halfords.com/webapp/wcs/stores/servlet/product_storeId_10001_catalogId_10151_productId_871069_langId_-1_categoryId_226211 [accessed May 2014]
5. UK Breathalysers website. Draeger Police 6510. http://www.ukbreathalysers.com/2_alcohol/test_others/prob/6510.html [accessed May 2014].
6. Gibb KA, Yee AS, Johnston CC, et al. Accuracy and usefulness of a breath alcohol analyzer. *Ann Emerg Med* 1984;13:516-20.
7. Road Traffic Act 1988 Chapter 52 Part I Sections 4-11. Motor vehicles: drink and drugs. Ministry of Justice, UK http://www.legislation.gov.uk/ukpga/1988/52/part/I/crossheading/motor-vehicles-drink-and-drugs [accessed May 2014].
8. Dräger Alcotest 6510 product specification 2012.
9. McCance RA, Widdowson, EM. McCance and Widdowson's The Composition of Foods. 6th edition compiled by the Food Standards Agency and Institute of Food Research, Cambridge, UK: Royal Society of Chemisty 2002.
10. Newcombe RG, Altman, DG. Proportions and their differences. In: Altman DG, Bryant TN, Gardner MJ, eds. Statistics With Confidence. 2nd edition ed. London: BMJ Books 2000:46-54.
11. Jones AW, Andersson L. Comparison of ethanol concentrations in venous blood and end-expired breath during a controlled drinking study. *Forensic Sci Int* 2003;132:18-25.

12. Parliamentary Under-Secretary of State, Home Office. The Preliminary Breath Test Device Approval 2005. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/117005/draeger-lion-envitec-alcoquant.pdf [accessed August 2014].
13. Home Office and Forensic Science Service. Breath Alcohol Screening Devices: A guide to type-approval procedures for breath alcohol screening devices used for law enforcement in Great Britain (Version 1.0, 30/06/2004). https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/117043/breathalcoholscreening.pdf [accessed August 2014].
14. Clapp JD, Min JW, Shillington AM, et al. Person and environment predictors of blood alcohol concentrations: a multi-level study of college parties. *Alcohol Clin Exp Res* 2008;32:100-7.
15. Bond J, Ye Y, Cherpitel CJ, et al. The relationship between self-reported drinking and BAC level in emergency room injury cases: is it a straight line? *Alcohol Clin Exp Res* 2010;34:1118-25.
16. Kraus CL, Salazar NC, Mitchell JR, et al. Inconsistencies between actual and estimated blood alcohol concentrations in a field study of college students: do students really know how much they drink? *Alcohol Clin Exp Res* 2005;29:1672-6.
17. Olson KN, Smith SW, Kloss JS, et al. Relationship between blood alcohol concentration and observable symptoms of intoxication in patients presenting to an emergency department. *Alcohol Alcohol* 2013;48:386-9.

Breathalyser	Total participants	Correctly positive by reference standard	Correctly negative by reference standard	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)
Alcosense Elite multi-use	205	34/38	107/167	89.5 (75.9 to 95.8)	64.1 (56.6 to 71.0)	36.2 (27.2 to 46.3)	96.4 (91.1 to 98.6)
Dräger Alco-Check single-use	108	18/19	45/89	94.7 (75.4 to 99.1)	50.6 (40.4 to 60.7)	29.0 (19.2 to 41.3)	97.8 (88.7 to 99.6)
Alcosense Single single-use	100	5/19	79/81	26.3 (11.8 to 48.8)	97.5 (91.4 to 99.3)	71.4 (35.9 to 91.8)	85.0 (76.3 to 90.8)

Table 1: Diagnostic accuracy of index breathalysers compared to reference police breathalyser, using researcher interpretation of single-use breathalysers. 95% confidence intervals shown in parentheses.

Threshold of self-reported alcohol consumption (UK units)	Total participants	Correctly positive by reference standard	Correctly negative by reference standard	Sensitivity (%)	Specificity (%)
1	207	38/38	9/169	100.0 (90.8 to 100.0)	5.3 (2.8 to 9.8)
2	207	38/38	23/169	100.0 (90.8 to 100.0)	13.6 (9.2 to 19.6)
3	207	38/38	52/169	100.0 (90.8 to 100.0)	30.8 (24.3 to 38.1)
4	207	38/38	71/169	100.0 (90.8 to 100.0)	42.0 (34.8 to 49.5)
5	207	37/38	84/169	97.4 (86.5 to 99.5)	49.7 (42.3 to 57.2)
6	207	35/38	105/169	92.1 (79.2 to 97.3)	62.1 (54.6 to 69.1)
7	207	34/38	117/169	89.5 (75.9 to 95.8)	69.2 (61.9 to 75.7)
8	207	29/38	130/169	76.3 (60.8 to 87.0)	76.9 (70.0 to 82.6)
9	207	26/38	142/169	68.4 (52.5 to 80.9)	84.0 (77.7 to 88.8)
10	207	20/38	148/169	52.6 (37.3 to 67.5)	87.6 (81.8 to 91.7)

Table 2: Diagnostic accuracy of self-reported alcohol consumption compared to reference police breathalyser. 1 UK unit is equivalent to 10ml or 8g alcohol (N.B. total 207 partcipants due to missing data on alcohol consumption for one participant). 95% confidence intervals shown in parentheses.

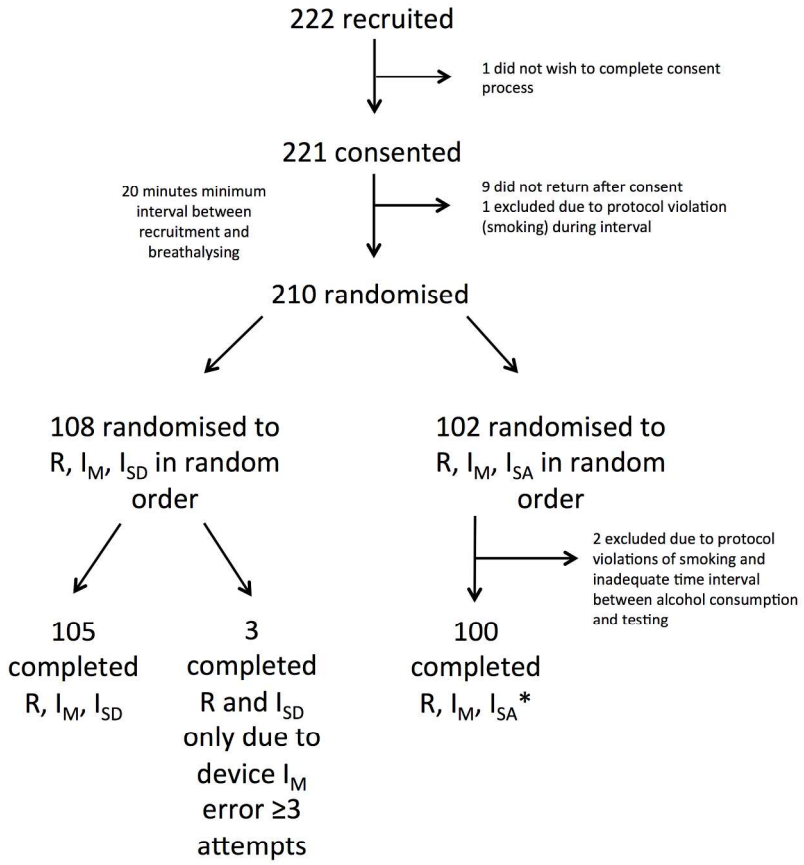
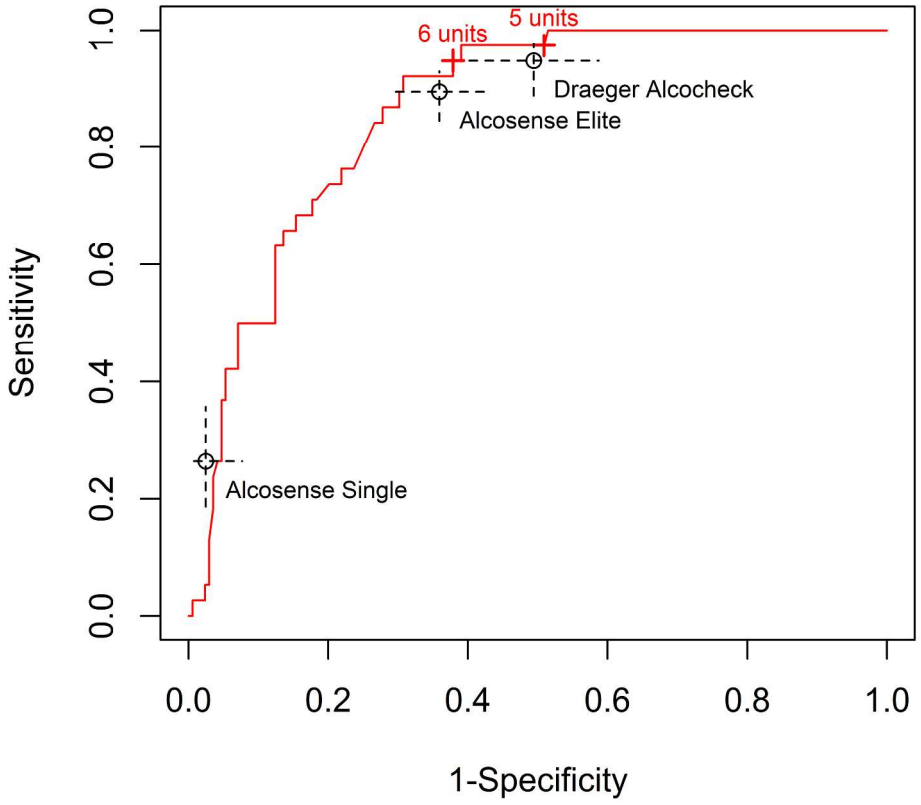


Figure 1: Study flow chart. All participants (except exclusions detailed in figure) were tested with the Dräger Alcotest 6510 (R), the Alcosense Elite (IM), and one of either the Dräger Alco-Check (ISD) or the Alcosense Single (ISA). The order of undertaking each breathalyser, and the selection of ISD or ISA was determined by randomisation. *3 participants left before analysis therefore results for participant estimation of ISA are only available for 97 participants.
190x254mm (300 x 300 DPI)



ROC curve of self-reported alcohol consumption in UK units, with comparative sensitivity and specificity for breathalysers tested.
127x127mm (600 x 600 DPI)

Web appendix: full 2x2 table data and diagnostic accuracy results for researcher and participant estimation and sensitivity analyses

Main analysis

Alcosense Elite

		Reference		
		Positive	Negative	Total
Index	Positive	34	60	94
	Negative	4	107	111
	Total	38	167	205

		Lower limit CI	*Upper limit CI*
Sensitivity	89.47%	*75.87%*	*95.83%*
Specificity	64.07%	*56.55%*	*70.96%*
PPV	36.17%	*27.18%*	*46.25%*
NPV	96.40%	*91.10%*	*98.59%*

Dräger Alco-Check

		Reference		
		Positive	Negative	Total
Index	Positive	18	44	62
	Negative	1	45	46
	Total	19	89	108

		Lower limit CI	*Upper limit CI*
Sensitivity	94.74%	*75.36%*	*99.06%*
Specificity	50.56%	*40.37%*	*60.71%*
PPV	29.03%	*19.22%*	*41.29%*
NPV	97.83%	*88.66%*	*99.62%*

Alcosense Single

		Reference		
		Positive	Negative	Total
Index	Positive	5	2	7
	Negative	14	79	93
	Total	19	81	100

		Lower limit CI	*Upper limit CI*
Sensitivity	26.32%	*11.81%*	*48.79%*
Specificity	97.53%	*91.44%*	*99.32%*
PPV	71.43%	*35.89%*	*91.78%*
NPV	84.95%	*76.30%*	*90.82%*

Participant estimation

Alcosense Elite

Not applicable as no participant estimation
(i.e. same as main analysis)

Dräger Alco-Check

		Reference		
		Positive	Negative	Total
Index	Positive	18	42	60
	Negative	1	47	48
	Total	19	89	108

		Lower limit CI	*Upper limit CI*
Sensitivity	94.74%	*75.36%*	*99.06%*
Specificity	52.81%	*42.54%*	*62.85%*
PPV	30.00%	*19.90%*	*42.51%*
NPV	97.92%	*89.10%*	*99.63%*

Alcosense Single

		Reference		
		Positive	Negative	Total
Index	Positive	3	4	7
	Negative	15	75	90
	Total	18	79	97

		Lower limit CI	*Upper limit CI*
Sensitivity	16.67%	*11.81%*	*48.79%*
Specificity	94.94%	*91.44%*	*99.32%*
PPV	42.86%	*35.89%*	*91.78%*
NPV	83.33%	*76.30%*	*90.82%*

Participant estimation - sensitivity analysis excluding colour blind person (A/168)

Alcosense Elite

Not applicable as no participant estimation
(i.e. same as main analysis)

Dräger Alco-Check

		Reference		
		Positive	Negative	Total
Index	Positive	18	41	59
	Negative	1	47	48
	Total	19	88	107

		Lower limit CI	*Upper limit CI*
Sensitivity	94.74%	*75.36%*	*99.06%*
Specificity	53.41%	*43.06%*	*63.47%*
PPV	30.51%	*20.25%*	*43.15%*
NPV	97.92%	*89.10%*	*99.63%*

Alcosense Single

Not applicable as colour blind person not in this group
(i.e. same as participant estimation above)

Researcher estimation - sensitivity analysis excluding suspect protocol violation (A/111)

Alcosense Elite

		Reference		
		Positive	Negative	Total
Index	Positive	34	60	94
	Negative	3	107	110
	Total	37	167	204

		Lower limit CI	*Upper limit CI*
Sensitivity	91.89%	*78.70%*	*97.20%*
Specificity	64.07%	*56.55%*	*70.96%*
PPV	36.17%	*27.18%*	*46.25%*
NPV	97.27%	*92.29%*	*99.07%*

Dräger Alco-Check

Not applicable as protocol violation person not in this group
(i.e. same as main analysis)

Alcosense Single

		Reference		
		Positive	Negative	Total
Index	Positive	5	2	7
	Negative	13	79	92
	Total	18	81	99

		Lower limit CI	*Upper limit CI*
Sensitivity	27.78%	*12.50%*	*50.87%*
Specificity	97.53%	*91.44%*	*99.32%*
PPV	71.43%	*35.89%*	*91.78%*
NPV	85.87%	*77.31%*	*91.55%*

Researcher estimation - sensitivity analysis excluding potential faults with devices (A/165 and A/179)

Alcosense Elite

Not applicable as no potential errors with Alcosense Elite
(i.e. same as main analysis)

Dräger Alco-Check

		Reference		
		Positive	Negative	Total
Index	Positive	18	44	62
	Negative	0	45	45
	Total	18	89	107

		Lower limit CI	*Upper limit CI*
Sensitivity	100.00%	*82.41%*	*100.00%*
Specificity	50.56%	*40.37%*	*60.71%*
PPV	29.03%	*19.22%*	*41.29%*
NPV	100.00%	*92.13%*	*100.00%*

Alcosense Single

		Reference		
		Positive	Negative	Total
Index	Positive	5	2	7
	Negative	14	78	92
	Total	19	80	99

		Lower limit CI	*Upper limit CI*
Sensitivity	26.32%	*11.81%*	*48.79%*
Specificity	97.50%	*91.34%*	*99.31%*
PPV	71.43%	*35.89%*	*91.78%*
NPV	84.78%	*76.06%*	*90.71%*

Table 1. STARD checklist for the reporting of studies of diagnostic accuracy.

Section and Topic	Item #		On page #
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	1
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	4
METHODS		Describe	
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where the data were collected.	5
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	5
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4? If not, specify how participants were further selected.	5
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	5
Test methods	7	The reference standard and its rationale.	
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	5
	9	Definition of and rationale for the units, cutoffs and/or categories of the results of the index tests and the reference standard.	6
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard.	5
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	5/6
Statistical methods	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	6
	13	Methods for calculating test reproducibility, if done.	N/a
RESULTS		Report	
Participants	14	When study was done, including beginning and ending dates of recruitment.	5
	15	Clinical and demographic characteristics of the study population (e.g. age, sex, spectrum of presenting symptoms, comorbidity, current treatments, recruitment centers).	7
	16	The number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test (a flow diagram is strongly recommended).	7/ Figure 1
Test results	17	Time interval from the index tests to the reference standard, and any treatment administered between.	6
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	7
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	7/ Table 1/ Web appendix
	20	Any adverse events from performing the index tests or the reference standard.	7
Estimates	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	7/Table 1/ Web appendix
	22	How indeterminate results, missing responses and outliers of the index tests were handled.	6/7/ Web appendix
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	7/ Web appendix
	24	Estimates of test reproducibility, if done.	N/a
DISCUSSION	25	Discuss the clinical applicability of the study findings.	8-10

BMJ Open

Diagnostic accuracy study of three alcohol breathalysers marketed for sale to the public

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-005811.R3
Article Type:	Research
Date Submitted by the Author:	14-Dec-2014
Complete List of Authors:	Ashdown, Helen; University of Oxford, Nuffield Department of Primary Care Health Sciences Fleming, Susannah; University of Oxford, Nuffield Department of Primary Care Health Sciences Spencer, Elizabeth; University of Oxford, Primary Care Health Sciences Thompson, Matthew; University of Oxford, Primary Care Health Sciences Stevens, Richard; University of Oxford, Department of Primary Health Care
Primary Subject Heading:	Diagnostics
Secondary Subject Heading:	Health policy, Addiction
Keywords:	BIOTECHNOLOGY & BIOINFORMATICS, PUBLIC HEALTH, TOXICOLOGY, PREVENTIVE MEDICINE

Diagnostic accuracy study of three alcohol breathalysers marketed for sale to the public

Helen F Ashdown[1], Susannah Fleming[1], Elizabeth A Spencer[1], Matthew J Thompson[1], Richard J Stevens[1]

[1] Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, United Kingdom

Helen F Ashdown
Clinical Researcher in General Practice

Susannah Fleming
Research Fellow

Elizabeth A Spencer
Nutritional Epidemiologist

Matthew J Thompson
Researcher

Richard J Stevens
University Research Lecturer in Medical Statistics

Correspondence to:
HF Ashdown
helen.ashdown@phc.ox.ac.uk

Word count (main text): 3119

ABSTRACT

Objectives

To assess the diagnostic accuracy of 3 personal breathalyser devices available for sale to the public marketed to test safety to drive after drinking alcohol

Design

Prospective comparative diagnostic accuracy study comparing two single-use breathalysers and one digital multi-use breathalyser (index tests) to a police breathalyser (reference test).

Setting

Establishments licensed to serve alcohol in a UK city

Participants

Of 222 participants recruited, 208 were included in the main analysis. Participants were eligible if they were 18 years old or over, had consumed alcohol and were not intending to drive within the following six hours.

Outcome measures

Sensitivity and specificity of the breathalysers for the detection of being at or over the UK legal driving limit (35 micrograms/100ml breath alcohol concentration).

Results

18% of participants (38/208) were at or over the UK driving limit according to the police breathalyser. The digital multi-use breathalyser had a sensitivity of 89.5% (95% CI 75.9 to 95.8%) and a specificity of 64.1% (95% CI 56.6 to 71.0%). The single-use breathalysers had a sensitivity of 94.7% (95% CI 75.4 to 99.1%) and 26.3% (95% CI 11.8 to 48.8%), and a specificity of 50.6% (95% CI 40.4 to 60.7%) and 97.5% (95% CI 91.4 to 99.3%) respectively. Self-reported alcohol consumption threshold of 5 UK units or fewer had a higher sensitivity than all personal breathalysers.

Conclusions

One alcohol breathalyser had sensitivity of 26%, corresponding to false reassurance for approximately one person in four who is over the limit by the reference standard, at least on the evening of drinking alcohol. The other devices tested had 90% sensitivity or higher. All estimates were subject to uncertainty. There is no clearly defined minimum sensitivity for this safety-critical application. We conclude that current regulatory frameworks do not ensure high sensitivity for these devices marketed to consumers for a decision with potentially catastrophic consequences.

Strengths and limitations of this study

- Personal breathalysers are available for sale to the public in pharmacies for assessing safety to drive following consuming alcohol, some very inexpensively, and in some jurisdictions is now promoted by law.
- Accuracy of personal breathalyser devices has never previously been studied.
- This study tested diagnostic accuracy of these devices in a real-life setting including participant estimation of readings.
- Limitations include the uncontrolled environment of public houses and bars, use of a pragmatic reference standard, and wide confidence intervals for some results due to low prevalence of those over the driving limit.
- However, our conclusions for the worst performing device are robust even against this uncertainty, since even the upper confidence limit for sensitivity of the worst performing device (48.8%) would still mean that about one in two individuals would be falsely reassured and this could lead to potentially dangerous driving decisions.

INTRODUCTION

Road traffic collisions (RTCs) are the second leading cause of death worldwide amongst people aged 5-29 years, estimated at 1.2 million deaths each year, and forecast to rise further by 2020.[1] Consumption of alcohol is an important factor influencing the likelihood and severity of RTCs, and has been found to be a causal factor in 17-40% of RTCs worldwide.[2] The risk of an RTC increases rapidly with increasing blood alcohol concentration, with relative risk rising significantly beyond a blood alcohol concentration (BAC) of 0.04g alcohol per 100ml blood (g/dl) to reach a relative risk of approximately 5 at 0.10g/dl.[2] Introduction of maximum legal BAC limits for driving, which vary from 0.02g/dl (e.g. Russia) to 0.15g/dl (e.g. Uganda), has been effective in reducing alcohol-related injuries and deaths.[1]

Alcohol breathalysers, which have long been available to law enforcement agencies, are now marketed direct to consumers, for example in some UK pharmacies and motoring stores, to test safety to drive following drinking alcohol, including the morning after. In July 2012, it became a legal requirement for drivers in France to carry a personal breathalyser at all times.[3] Devices marketed to consumers are order of magnitudes lower in cost than those intended for law enforcement: for example at the time of writing the Alcosense Single is more than 300 times cheaper than the Dräger 6510 Home Office certified police breathalyser.[4, 5]

The theoretical accuracy of breath alcohol measurement as a surrogate for blood alcohol measurement has been well-studied,[6] and in the UK breath alcohol forms part of the prescribed legal limit, along with blood and urine alcohol concentration.[7] However to our knowledge the accuracy of devices currently marketed to the consumer/motorist has not been studied. Many such devices carry regulatory approvals such as the Conformité Européenne (CE), French NF certification and British Standard (BSI Kitemark) but in general such marks are statements of engineering quality rather than diagnostic accuracy. We therefore aimed to assess the diagnostic accuracy of personal breathalysers compared to a police breathalyser for detection of being at or over the UK legal driving limit in a real world situation representing a possible application for such devices.

METHODS

Index tests and reference standard

We selected for study as index tests the Alcosense Single, the UK counterpart of an NF-approved device widely sold for motorists in France; the Dräger Alco-Check, as a comparable single-use device from a competing manufacturer; and the Alcosense Elite, as an example of a digital multi-use device readily available from pharmacies, high street and online stores (Boots, Halfords, Amazon and others). We selected as reference standard the Dräger Alcotest 6510 device. This has Home Office approval and is standard issue to UK police for use as an initial test at the roadside,[5] and is approved in the US as an evidential breath testing device (National Highway Traffic Safety Administration Standard 49 FR 48854) meaning readings can be used as evidence for prosecution of drink driving offences in US courts. Manufacturer information states measurement precision as ±0.008mg/L or 1.7% of measurement value.[8] The Dräger Alcotest 6510 and Alcosense Elite devices come with instruction manuals including details of breathalyser operating technique and result interpretation. The single-use devices contain information on their packaging and an enclosed sheet of paper including details of breathalyser operating technique and result interpretation. These manuals and leaflets recommended at least 15 minutes (Dräger Alcocheck), at least 20 minutes (Dräger 6510 and Alcosense Single), or at least 30 minutes (Alcosense Elite) elapse between alcohol consumption and use. All personal breathalyser instructions clearly state that any amount of alcohol even below the limit can impair driving ability.

Study participants

To test the breathalysers in a real world situation, representing one possible application for personal breathalysers, we recruited participants from establishments licensed to serve alcohol in the city centre of Oxford, UK, including college bars and public houses. In the absence of prior data with which to estimate sample size, we decided to recruit 200 participants, which would allow us to report the prevalence of intoxication with small standard error (maximum 3.5%). Recruitment took place on evenings during the period from October 2012 to January 2013: 11 study evenings were required to reach a sample size greater than 200. Participants were eligible if they were 18 years or over, had consumed alcohol, and were not intending to drive within the following six hours. We excluded

potential drivers to ensure that no participant could inadvertently incriminate themselves by taking part in the study.

Recruitment and consent

Individuals in these establishments were informally approached by a member of the research team and given preliminary information about the study. Potentially eligible interested participants were then asked to sit with the research team, at tables reserved within the same premises, where they received a full description of the study, were asked to read the study literature, and given an opportunity to ask further questions. Eligibility was checked and written consent taken. Participants were given a card with contact details to use in the event of withdrawing consent subsequently (for example, the next day): in the event, no participants used this option to withdraw consent.

Study procedures

The research team consisted of between two and four individuals, who followed the reference breathalyser manufacturers' written instructions in directing their use: a minimum of 20 minutes was enforced between recruitment and using the breathalyser devices, and participants were asked not to drink further alcohol, smoke, use mouthwash, or drink fruit juice during this period. Participants were provided with water and asked to take at least one sip in order to clear any residual alcohol from the upper airway. During this 20 minute period basic demographic details (age and sex) and reported alcoholic drinks consumption during the preceding 12 hours were recorded by the researchers. Participants were not required to remain under observation during the 20 minute period.

The 20 minute waiting period did not comply with the manufacturers' instructions for one of the devices (Alcosense Elite). However, it was felt that a longer gap between recruitment and testing would result in high attrition rates, and increased potential for protocol violation by participants (e.g. by smoking or drinking liquids other than water). We therefore compromised on a waiting time of 20 minutes based on the instructions for the reference device, and the fact that this complied with manufacturer instructions for the other two index devices.

Participants then used three breathalysers, at intervals of at least one minute, at the study tables under the supervision of a research team member. Each participant used, in random

order, the reference standard, the Alcosense Elite multi-use and a randomly selected single use breathalyser. Randomisation was carried out in advance by study number in random permuted blocks.

A member of the research team recorded digital outputs of the multi-use and reference breathalysers. The single-use breathalysers require subjective assessment of colour of crystals after use: we recorded assessments by both researcher (primary outcome) and participant (secondary outcome). Participants were blinded to the researcher assessment and to the reference result.

Statistical analysis

We calculated sensitivity, specificity, positive and negative predictive values of the breathalysers for the outcome of being at or over the current UK legal driving limit[7] of 35 micrograms/100ml or 0.8‰BAC according to the reference standard. Self-reported alcohol consumption, converted to UK alcohol units (1 unit = 8g alcohol) using nutritional tables,[9] was also assessed for diagnostic accuracy of being at or over the UK legal driving limit, and an ROC analysis undertaken to assess different unit thresholds. Statistical calculations were carried out using standard methods[10] and the ROC analysis using Stata (Release 11). Minor protocol violations were discussed amongst the research team, with sensitivity analyses performed to determine whether there was any difference in overall results due to inclusion or exclusion of these. Participants with missing data (for example sex, self-reported alcohol consumption or participant estimation of result) were included except for analyses involving the missing component itself.

RESULTS

A total of 208 participants were included in the main analysis (see Figure 1 flowchart), of whom 148/207 (71.5%) were male and with a median age of 20. Participants reported having consumed a median of 6 UK units of alcohol (range 1 to 25), equivalent to a median of 46g (range 8 to 204) alcohol. 108 participants were tested with the Dräger Alco-check single-use breathalyser, and 100 with the Alcosense Single single-use breathalyser.

38 participants (18.3%, 95% CI 11.7 to 27.4%) were at or above the current UK driving limit of 35 μg/100ml according to the reference breathalyser. Table 1 compares performance of the three index breathalysers at detecting those at or over the UK limit. Compared to the reference breathalyser, the Alcosense Elite multi-use breathalyser had a sensitivity of 89.5% (95% CI 75.9 to 95.8%), the Dräger Alco-Check single-use breathalyser had a sensitivity of 94.7% (95% CI 75.4 to 99.1%) and the Alcosense Single breathalyser had a sensitivity of 26.3% (95% CI 11.8 to 48.8%). When analyses were repeated using the participant's interpretation of colour change in the single-use breathalysers instead of researcher's interpretation, sensitivity was 94.7% (95% CI 75.4 to 99.1%) for the Dräger Alco-Check (i.e. identical to researcher estimation) and 16.7% (95% CI 11.8 to 48.8%) for the Alcosense Single (see Appendix 1 supplementary material for full data for participant estimation).

We conducted three sensitivity analyses in turn (i) excluding two results where the participants were suspected to have incorrectly used the device, (ii) a colour-blind participant who may have had difficulty interpreting the colour change of crystals, and (iii) a participant who was suspected to have violated the protocol (consumed alcohol between use of the three devices). In general these analyses showed minimal difference to results overall (see Appendix 1) with the exception that the sensitivity of the Dräger Alco-Check increased to 100% (95% CI 83.2 to 100.0%) in the latter sensitivity analysis. There were no reported adverse events from using any of the breathalyser devices.

Because there is no single standard threshold for driving decisions based on self-estimated alcohol consumption, we calculated sensitivity and specificity for all alcohol unit thresholds (Table 2) and plotted them in ROC space (Figure 2).

DISCUSSION

We have shown that breathalysers available for sale to the public for personal use vary considerably in their performance in detecting being at or over a legal driving limit during the period directly after drinking. Two of the devices tested, the Alcosense Elite digital multi-use breathalyser and Dräger Alco-Check single-use breathalyser had a sensitivity of approximately 90 and 95 per cent respectively in the main analyses. However, even a sensitivity of 95% means that approximately 1 in 20 people over the driving limit would be falsely reassured by these tests. We question whether even this would be sufficient sensitivity to assess safety to drive. The third device, the Alcosense Single single-use breathalyser, had a sensitivity of only 26 per cent, meaning that only approximately one in four individuals over the legal limit would be identified by this device. Participants (rather than researchers) interpreting results, which is what would occur in real life, reduced sensitivity further to only 17 per cent. This device has a correspondingly high specificity, but specificity is not the safety-critical aspect of performance of a device assessing safety to drive. Surprisingly, we found that self-reported consumption of alcohol was a more sensitive test for being at or over the legal driving limit than the three breathalysers tested for up to 5 UK units of alcohol consumed. None of the devices outperformed simple recall of amount of alcohol consumed up to 5 UK units of alcohol.

Strengths of our study include testing participants in a "real world" environment including assessment of participants' estimation of breathalyser readings, which enables generalizability to everyday life. However, this pragmatic approach also brings limitations in that the setting in college bars and public houses could not be rigorously controlled. It is possible, for example, that poor lighting, or unknown protocol violation by subjects who could not always be perfectly monitored in this busy environment, could have introduced some inaccuracies. Operating the three breathalysers in close succession and randomisation of order of use of the breathalysers should have helped reduce the risk of such effects causing bias in the overall results. There is also the possibility that ambient alcohol vapour in the environment may have resulted in excess false positives. As discussed above, we may have underestimated the sensitivity of the Dräger Alco-Check because of a suspected protocol violation. Three participants inadvertently took the breathalysers in a different order to that planned: we decided to include these participants because the aim was to have an overall

variety in breathalyser order and this was unlikely to have any impact for such a small number of participants. The use of a 20 minute rather than 30 minute minimum time after drinking alcohol was not adherent to the manufacturer's instructions for the Alcosense Elite, and so it is possible that this may have affected results for this breathalyser. Hyperventilation immediately prior to a breath sample reduces breath alcohol concentration [11] and breath holding increases it.[12,13] We recorded a minimum amount of information about participants to facilitate recruitment and monitoring breathing would not have been possible in this deliberately real life environment. However, we followed the instructions in the Alcosense Elite manual to "wait until you are breathing normally again" by ensuring that participants were relaxed and ensuring at least a minute between breathalysers for any recovery required. Participants were typically seated while completing the consent process and initial data collection before breathalysing took place. Randomising the order of breathalysing would further reduce any bias in the overall results due to temporal effects such as hyperventilation before the first test. The high solubility of alcohol means that it is thought to be deposited in exhaled air largely from the proximal airway conducting system, and breath alcohol concentration is not therefore reliant on alveolar equilibration.[14] Our methodology of allowing one minute between each test was designed to allow satisfactory measurement of breath alcohol concentration.

Participants were blinded to their result on the reference breathalyser until after completing estimation of colour change, however researchers were not and it is possible that this could have introduced some bias. We met our planned recruitment target, but the prevalence of those over the legal limit by the reference standard means that our confidence intervals are fairly wide. Even using the upper confidence limit for sensitivity, however, the worst-performing breathalyser would still have sensitivity under 50 per cent. Our sample of participants obtained from colleges and pubs may not be representative of the population who purchase personal breathalysers, particularly in age group and quantity of alcohol consumed and socio-demographics; although this might affect sensitivity of self-reported alcohol consumption, it should not impact significantly on the overall accuracy of the breathalyser devices. We tested the index devices on the night of drinking and in a bar environment and the generalizability to other uses such as "morning-after" use may be unclear; blood and breath alcohol concentrations decline with time after drinking alcohol at similar rates and maintain high correlation,[15] but it is possible that diagnostic accuracy of personal breathalysers may differ at lower alcohol concentrations.

Use of blood alcohol as a gold standard was not possible in this pragmatic, field study. We chose as our reference standard the Dräger Alcotest 6510 which has passed a home office testing protocol requiring error less than 10% in all readings [16,17] and is also an evidential device for criminal prosecutions in the US. The devices currently approved for evidential use in the UK are not portable and were therefore not suitable for our study setting. For convenience we selected index devices easily available in the UK for testing, however the same devices calibrated for different legal limits are sold outside of the UK, and other devices sold elsewhere use similar technology and are similarly priced. Therefore, while we cannot directly apply our results to other countries, we would anticipate similar findings elsewhere.

We have not found other studies testing the accuracy of personal breathalysers. Studies in the US of college student parties have found a mean BAC of 0.077% (standard deviation 0.063%) (for context, the UK driving limit is 0.08%).[18] A Canadian study found a linear relationship between self-reported alcohol consumption and BAC in Emergency Room attendees, up to 7 drinks.[19] However, a US study of college students which compared estimated BAC to measured BAC, found that students tended to over-estimate their levels of consumption when surveyed in the midst of a night of drinking.[20] We did not attempt to convert self-reported alcohol consumption to BAC and only recorded total quantity consumed, which would have had differential effects between individuals dependent on weight, sex and the time over which the drinks were consumed. However, despite the variation these factors would introduce into BAC, self-reported consumption up to 5 units was still a more sensitive test than the breathalysers tested, and BAC is known to correlate poorly with symptoms of intoxication.[21]

Our research suggests that at least some personal breathalysers available for sale to the public are not always sufficiently sensitive to test safety to drive after drinking alcohol, where use of inaccurate information from breathalysers thought to be accurate could have catastrophic safety implications for drivers. The fact that these devices are sold in well-established pharmacies including national chains does not guarantee sufficient accuracy for safe use. Medical and measurement devices may carry regulatory approvals such as CE- or NF-marking, but this does not appear to correlate with accuracy, and this raises wider questions over how this marking may be perceived by users. A derivative device of the worst-performing breathalyser in our study is widely sold for use in France as part of the new law

requiring breathalysers to be carried when driving, and has French NF-approval.[4] Although results from our study cannot be directly applied to the lower French driving limit of 0.05g/dl and a derivative device, they question the utility of the new law which on the one hand may improve public awareness of drink-driving in general, but also risks ill-informed driving decisions based on inaccurate results from a personal breathalyser. Replication of our results in other settings and with other breathalysers could further inform policy makers planning to introduce similar laws in other jurisdictions, and explore the characteristics of the population who purchase personal breathalysers and how they use the results obtained. Finally, our research raises worrying questions about the level of scrutiny that medical tests intended for sale to the general public undergo in Europe, and raises wider concerns about how diagnostic accuracy in particular is evaluated, and whether any further field evaluations are required for intended users, perceptions of accuracy of such devices and how use of such devices interacts with medical testing in other health care settings.

REFERENCES

1. Peden M, Scurfield R, Sleet D, et al. World report on road traffic injury prevention. Geneva: World Health Organization 2004.
2. Global Road Safety Partnership. Drinking and driving: a road safety manual for decision-makers and practitioners. Geneva: World Health Organization 2007.
3. Légifrance: Le Service Public pour la Diffusion du Droit: Décret n° 2012-284 du 28 février 2012 relatif à la possession obligatoire d'un éthylotest par le conducteur d'un véhicule terrestre à moteur http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000025417826&categorieLien=id [accessed May 2014].
4. Halfords store website. Alcosense Single breathalyser http://www.halfords.com/webapp/wcs/stores/servlet/product_storeId_10001_catalogId_10151_productId_871069_langId_-1_categoryId_226211 [accessed May 2014]
5. UK Breathalysers website. Draeger Police 6510. http://www.ukbreathalysers.com/2_alcohol/test_others/prob/6510.html [accessed May 2014].
6. Gibb KA, Yee AS, Johnston CC, et al. Accuracy and usefulness of a breath alcohol analyzer. *Ann Emerg Med* 1984;13:516-20.
7. Road Traffic Act 1988 Chapter 52 Part I Sections 4-11. Motor vehicles: drink and drugs. Ministry of Justice, UK http://www.legislation.gov.uk/ukpga/1988/52/part/I/crossheading/motor-vehicles-drink-and-drugs [accessed May 2014].
8. Dräger Alcotest 6510 product specification 2012.
9. McCance RA, Widdowson, EM. McCance and Widdowson's The Composition of Foods. 6th edition compiled by the Food Standards Agency and Institute of Food Research, Cambridge, UK: Royal Society of Chemisty 2002.
10. Newcombe RG, Altman, DG. Proportions and their differences. In: Altman DG, Bryant TN, Gardner MJ, eds. Statistics With Confidence. 2nd edition ed. London: BMJ Books 2000:46-54.
11. Ohlsson J, Ralph DD, Mandelkorn MA, et al. Accurate measurement of blood alcohol concentration with isothermal rebreathing. *J Stud Alcohol* 1990; 51:6-13.

12. Jones AW. How breathing technique can influence the results of breath-alcohol analysis. *Med Sci Law* 1982; 22:275-80.
13. Jones AW. Role of rebreathing in determination of the blood-breath ratio of expired ethanol. *J Appl Physiol Respir Environ Exerc Physiol* 1983; 4:1237-41.
14. Hlastala MP. The alcohol breath test – a review. *J Appl Physiol* 1998; 84:401-8.
15. Jones AW, Andersson L. Comparison of ethanol concentrations in venous blood and end-expired breath during a controlled drinking study. *Forensic Sci Int* 2003;132:18-25.
16. Parliamentary Under-Secretary of State, Home Office. The Preliminary Breath Test Device Approval 2005. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/117005/draeger-lion-envitec-alcoquant.pdf [accessed August 2014].
17. Home Office and Forensic Science Service. Breath Alcohol Screening Devices: A guide to type-approval procedures for breath alcohol screening devices used for law enforcement in Great Britain (Version 1.0, 30/06/2004). https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/117043/breathalcoholscreening.pdf [accessed August 2014].
18. Clapp JD, Min JW, Shillington AM, et al. Person and environment predictors of blood alcohol concentrations: a multi-level study of college parties. *Alcohol Clin Exp Res* 2008;32:100-7.
19. Bond J, Ye Y, Cherpitel CJ, et al. The relationship between self-reported drinking and BAC level in emergency room injury cases: is it a straight line? *Alcohol Clin Exp Res* 2010;34:1118-25.
20. Kraus CL, Salazar NC, Mitchell JR, et al. Inconsistencies between actual and estimated blood alcohol concentrations in a field study of college students: do students really know how much they drink? *Alcohol Clin Exp Res* 2005;29:1672-6.
21. Olson KN, Smith SW, Kloss JS, et al. Relationship between blood alcohol concentration and observable symptoms of intoxication in patients presenting to an emergency department. *Alcohol Alcohol* 2013;48:386-9.

Competing interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

The authors would like to thank the study participants, and those colleges of the University of Oxford and public houses in Oxford which permitted them to recruit on their premises; Michael Ashdown for assistance with legal literature; and David McCartney for additional advice on the final manuscript. We would also like to thank the reviewers of this article for their hard work and insightful comments.

Contributors

All authors contributed to study design, data collection, data interpretation and helped write the paper. RJS conceived the idea for the study. HFA, EAS and RJS analysed the data. HFA is guarantor. All authors approved the final version of the manuscript.

Ethical approval

The study was approved by the University of Oxford Medical Sciences Division Interdivisional Research Ethics Committee (reference MSD/IDREC/2011/13). All participants gave informed consent before taking part. No remuneration or incentives were provided to participants for taking part.

Funding

This paper summarises independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0407-10347). The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Sponsorship

The University of Oxford acted as study sponsor and had no role in study design, data collection, analysis or interpretation, writing of the paper or the decision to submit for publication. All authors were independent from funders and sponsors.

Data sharing

All authors had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. Full data are available from the corresponding author on request. Consent for data sharing was not obtained but the presented data are anonymised and risk of identification is low.

Transparency declaration

The lead author and the manuscript's guarantor affirm 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 any discrepancies from the study as planned have been explained.

Breathalyser	Total participants	Correctly positive by reference standard	Correctly negative by reference standard	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)
Alcosense Elite multi-use	205	34/38	107/167	89.5 (75.9 to 95.8)	64.1 (56.6 to 71.0)	36.2 (27.2 to 46.3)	96.4 (91.1 to 98.6)
Dräger Alco-Check single-use	108	18/19	45/89	94.7 (75.4 to 99.1)	50.6 (40.4 to 60.7)	29.0 (19.2 to 41.3)	97.8 (88.7 to 99.6)
Alcosense Single single-use	100	5/19	79/81	26.3 (11.8 to 48.8)	97.5 (91.4 to 99.3)	71.4 (35.9 to 91.8)	85.0 (76.3 to 90.8)

Table 1: Diagnostic accuracy of index breathalysers compared to reference police breathalyser, using researcher interpretation of single-use breathalysers. 95% confidence intervals shown in parentheses.

Threshold of self-reported alcohol consumption (UK units)	Total participants	Correctly positive by reference standard	Correctly negative by reference standard	Sensitivity (%)	Specificity (%)
1	207	38/38	9/169	100.0 (90.8 to 100.0)	5.3 (2.8 to 9.8)
2	207	38/38	23/169	100.0 (90.8 to 100.0)	13.6 (9.2 to 19.6)
3	207	38/38	52/169	100.0 (90.8 to 100.0)	30.8 (24.3 to 38.1)
4	207	38/38	71/169	100.0 (90.8 to 100.0)	42.0 (34.8 to 49.5)
5	207	37/38	84/169	97.4 (86.5 to 99.5)	49.7 (42.3 to 57.2)
6	207	35/38	105/169	92.1 (79.2 to 97.3)	62.1 (54.6 to 69.1)
7	207	34/38	117/169	89.5 (75.9 to 95.8)	69.2 (61.9 to 75.7)
8	207	29/38	130/169	76.3 (60.8 to 87.0)	76.9 (70.0 to 82.6)
9	207	26/38	142/169	68.4 (52.5 to 80.9)	84.0 (77.7 to 88.8)
10	207	20/38	148/169	52.6 (37.3 to 67.5)	87.6 (81.8 to 91.7)

Table 2: Diagnostic accuracy of self-reported alcohol consumption compared to reference police breathalyser. 1 UK unit is equivalent to 10ml or 8g alcohol (N.B. total 207 partcipants due to missing data on alcohol consumption for one participant). 95% confidence intervals shown in parentheses.

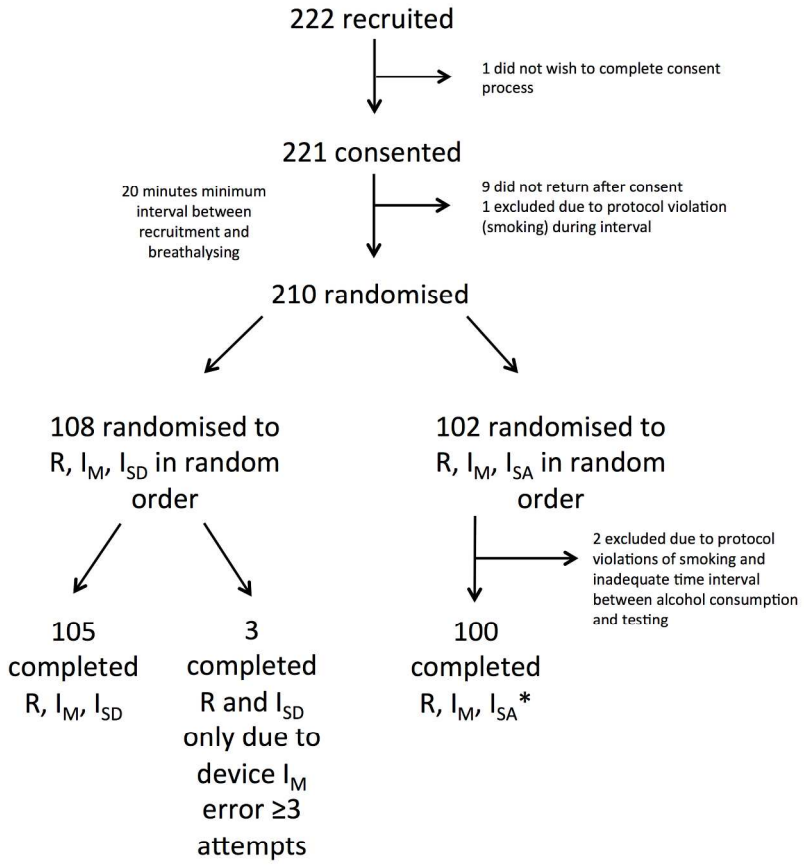
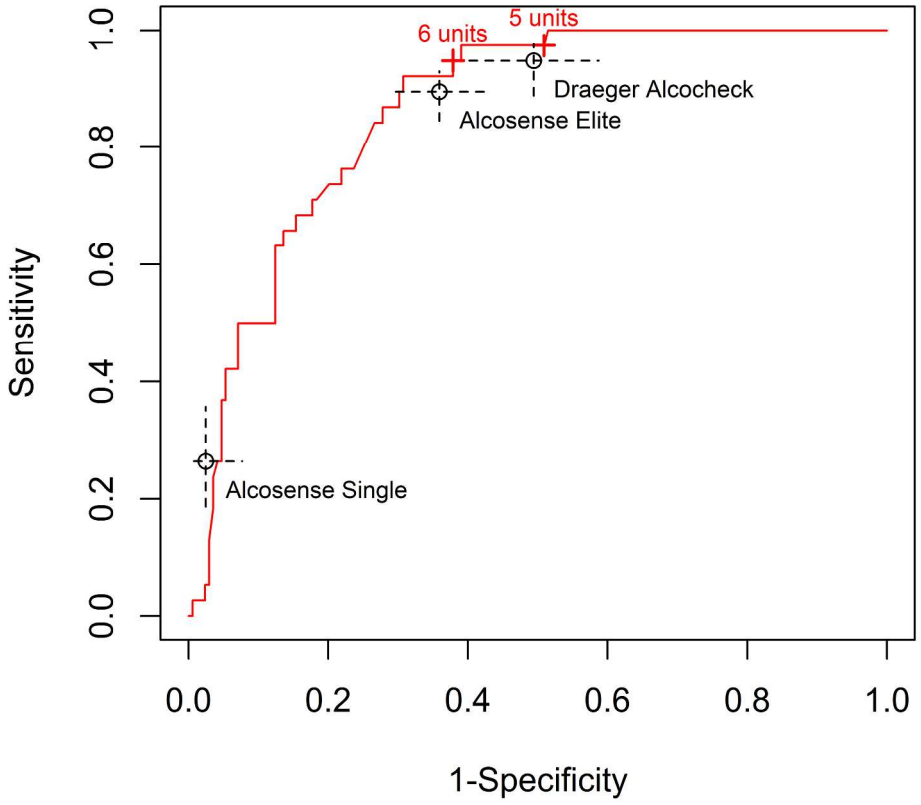


Figure 1: Study flow chart. All participants (except exclusions detailed in figure) were tested with the Dräger Alcotest 6510 (R), the Alcosense Elite (IM), and one of either the Dräger Alco-Check (ISD) or the Alcosense Single (ISA). The order of undertaking each breathalyser, and the selection of ISD or ISA was determined by randomisation. *3 participants left before analysis therefore results for participant estimation of ISA are only available for 97 participants.
190x254mm (300 x 300 DPI)



ROC curve of self-reported alcohol consumption in UK units, with comparative sensitivity and specificity for breathalysers tested.
127x127mm (600 x 600 DPI)

Web appendix: full 2x2 table data and diagnostic accuracy results for researcher and participant estimation and sensitivity analyses

Main analysis

Alcosense Elite

		Reference		
		Positive	Negative	Total
Index	Positive	34	60	94
	Negative	4	107	111
	Total	38	167	205

		Lower limit CI	*Upper limit CI*
Sensitivity	89.47%	*75.87%*	*95.83%*
Specificity	64.07%	*56.55%*	*70.96%*
PPV	36.17%	*27.18%*	*46.25%*
NPV	96.40%	*91.10%*	*98.59%*

Dräger Alco-Check

		Reference		
		Positive	Negative	Total
Index	Positive	18	44	62
	Negative	1	45	46
	Total	19	89	108

		Lower limit CI	*Upper limit CI*
Sensitivity	94.74%	*75.36%*	*99.06%*
Specificity	50.56%	*40.37%*	*60.71%*
PPV	29.03%	*19.22%*	*41.29%*
NPV	97.83%	*88.66%*	*99.62%*

Alcosense Single

		Reference		
		Positive	Negative	Total
Index	Positive	5	2	7
	Negative	14	79	93
	Total	19	81	100

		Lower limit CI	*Upper limit CI*
Sensitivity	26.32%	*11.81%*	*48.79%*
Specificity	97.53%	*91.44%*	*99.32%*
PPV	71.43%	*35.89%*	*91.78%*
NPV	84.95%	*76.30%*	*90.82%*

Participant estimation

Alcosense Elite

Not applicable as no participant estimation
(i.e. same as main analysis)

Dräger Alco-Check

		Reference		
		Positive	Negative	Total
Index	Positive	18	42	60
	Negative	1	47	48
	Total	19	89	108

		Lower limit CI	*Upper limit CI*
Sensitivity	94.74%	*75.36%*	*99.06%*
Specificity	52.81%	*42.54%*	*62.85%*
PPV	30.00%	*19.90%*	*42.51%*
NPV	97.92%	*89.10%*	*99.63%*

Alcosense Single

		Reference		
		Positive	Negative	Total
Index	Positive	3	4	7
	Negative	15	75	90
	Total	18	79	97

		Lower limit CI	*Upper limit CI*
Sensitivity	16.67%	*11.81%*	*48.79%*
Specificity	94.94%	*91.44%*	*99.32%*
PPV	42.86%	*35.89%*	*91.78%*
NPV	83.33%	*76.30%*	*90.82%*

Participant estimation - sensitivity analysis excluding colour blind person (A/168)

Alcosense Elite

Not applicable as no participant estimation
(i.e. same as main analysis)

Dräger Alco-Check

		Reference		
		Positive	Negative	Total
Index	Positive	18	41	59
	Negative	1	47	48
	Total	19	88	107

		Lower limit CI	*Upper limit CI*
Sensitivity	94.74%	*75.36%*	*99.06%*
Specificity	53.41%	*43.06%*	*63.47%*
PPV	30.51%	*20.25%*	*43.15%*
NPV	97.92%	*89.10%*	*99.63%*

Alcosense Single

Not applicable as colour blind person not in this group
(i.e. same as participant estimation above)

Researcher estimation - sensitivity analysis excluding suspect protocol violation (A/111)

Alcosense Elite

		Reference		
		Positive	Negative	Total
Index	Positive	34	60	94
	Negative	3	107	110
	Total	37	167	204

		Lower limit CI	*Upper limit CI*
Sensitivity	91.89%	*78.70%*	*97.20%*
Specificity	64.07%	*56.55%*	*70.96%*
PPV	36.17%	*27.18%*	*46.25%*
NPV	97.27%	*92.29%*	*99.07%*

Dräger Alco-Check

Not applicable as protocol violation person not in this group
(i.e. same as main analysis)

Alcosense Single

		Reference		
		Positive	Negative	Total
Index	Positive	5	2	7
	Negative	13	79	92
	Total	18	81	99

		Lower limit CI	*Upper limit CI*
Sensitivity	27.78%	*12.50%*	*50.87%*
Specificity	97.53%	*91.44%*	*99.32%*
PPV	71.43%	*35.89%*	*91.78%*
NPV	85.87%	*77.31%*	*91.55%*

Researcher estimation - sensitivity analysis excluding potential faults with devices (A/165 and A/179)

Alcosense Elite

Not applicable as no potential errors with Alcosense Elite
(i.e. same as main analysis)

Dräger Alco-Check

		Reference		
		Positive	Negative	Total
Index	Positive	18	44	62
	Negative	0	45	45
	Total	18	89	107

		Lower limit CI	*Upper limit CI*
Sensitivity	100.00%	*82.41%*	*100.00%*
Specificity	50.56%	*40.37%*	*60.71%*
PPV	29.03%	*19.22%*	*41.29%*
NPV	100.00%	*92.13%*	*100.00%*

Alcosense Single

		Reference		
		Positive	Negative	Total
Index	Positive	5	2	7
	Negative	14	78	92
	Total	19	80	99

		Lower limit CI	*Upper limit CI*
Sensitivity	26.32%	*11.81%*	*48.79%*
Specificity	97.50%	*91.34%*	*99.31%*
PPV	71.43%	*35.89%*	*91.78%*
NPV	84.78%	*76.06%*	*90.71%*

Diagnostic accuracy Study of three alcohol breathalyzers marketed for sale to the public

Aghdam HF et al

Table 1. STARD checklist for the reporting of studies of diagnostic accuracy.

Section and Topic	Item #		On page #
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	1
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	4
METHODS		Describe	
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where the data were collected.	5
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	5
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4? If not, specify how participants were further selected.	5
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	5
Test methods	7	The reference standard and its rationale.	
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	5
	9	Definition of and rationale for the units, cutoffs and/or categories of the results of the index tests and the reference standard.	6
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard.	5
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	5/6
Statistical methods	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	6
	13	Methods for calculating test reproducibility, if done.	N/a
RESULTS		Report	
Participants	14	When study was done, including beginning and ending dates of recruitment.	5
	15	Clinical and demographic characteristics of the study population (e.g. age, sex, spectrum of presenting symptoms, comorbidity, current treatments, recruitment centers).	7
	16	The number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test (a flow diagram is strongly recommended).	7/ Figure 1
Test results	17	Time interval from the index tests to the reference standard, and any treatment administered between.	6
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	7
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	7/ Table 1/ Web appendix
	20	Any adverse events from performing the index tests or the reference standard.	7
Estimates	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	7/Table 1/ Web appendix
	22	How indeterminate results, missing responses and outliers of the index tests were handled.	6/7/ Web appendix
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	7/ Web appendix
	24	Estimates of test reproducibility, if done.	N/a
DISCUSSION	25	Discuss the clinical applicability of the study findings.	8-10