Evaluation of symptom checkers for self diagnosis and triage: audit study

Hannah L Semigran,[1] Jeffrey A Linder,[2] Courtney Gidengil,[3 4] Ateev Mehrotra[1 5]

EDITORIAL by Wyatt

[1]Department of Health Care Policy, Harvard Medical School, Boston, MA 02115, USA
[2]Division of General Medicine and Primary Care, Brigham and Women's Hospital & Harvard Medical School, Boston, MA, USA
[3]Division of Infectious Diseases, Boston Children's Hospital & Harvard Medical School, Boston, MA, USA
[4]RAND Corporation, Boston, MA, USA
[5]Division of General Internal Medicine and Primary Care, Beth Israel Deaconess Medical Center, Boston, MA, USA
Correspondence to: A Mehrotra
mehrotra@hcp.med.harvard.edu

Cite this as: *BMJ* 2015;351:h3480
doi: 10.1136/bmj.h3480

This is a summary of a paper that was published on thebmj.com as *BMJ* 2015;351:h3480

thebmj.com
Watch the video at thebmj.com/content/351/bmj.h3480

STUDY QUESTION

What is the clinical accuracy of symptom checkers for diagnosis and triage?

SUMMARY ANSWER

Symptom checkers provided the correct diagnosis in one third of evaluations and suggested appropriate triage in approximately half of the evaluations. Although both outcomes varied by the severity of the condition, advice on triage was generally risk averse, suggesting users seek care for their conditions when medical attention was not necessary.

WHAT IS KNOWN AND WHAT THIS PAPER ADDS

Members of the public are increasingly turning to the internet to research their health concerns, and symptom checkers, which are offered by physicians' organisations, health plans, governments, and private companies, attempt to streamline and improve this process. This was the first large evaluation of the clinical performance of symptom checkers.

Selection criteria for samples

Using searches of Google and app stores, we identified 23 symptom checkers that were in English, were free, were publicly available, and focused on general medical advice. We used 45 standardised patient vignettes divided equally into three types of conditions: emergent care is required, non-emergent care is reasonable, and a medical visit is generally unnecessary and self care is sufficient.

Primary outcomes

The main outcome measures were whether the symptom checker listed the correct diagnosis first or within the first 20 potential diagnoses, and whether the symptom checker correctly recommended seeking emergent care, non-emergent care, or self care.

Accuracy of diagnosis decision and triage advice for all symptom checkers, stratified by severity of standardised patient (SP) evaluation. Values are percentages (95% confidence intervals)

	Listed correct diagnosis		Provided appropriate triage advice
SP evaluations	First	In top 20	
Overall	34 (31 to 37)	58 (55 to 62)	57 (52 to 61)
Type of SP evaluation:			
Emergent	24 (19 to 30)*	50 (44 to 56)*	80 (75 to 86)*
Non-emergent	38 (32 to 44)*	60 (54 to 66)*	55 (47 to 63)*
Self care	40 (34 to 47)*	65 (59 to 71)*	33 (26 to 40)*

*P<0.01 x^2 test evaluating whether diagnosis or triage was correct by severity of SP evaluation.

Main results and role of chance

The 23 identified symptom checkers were based in the United Kingdom, United States, Netherlands, and Poland: 11 provided both diagnoses and triage advice, eight provided only diagnoses, and four provided only triage advice. Performance was assessed on a total of 770 standardised patient evaluations for diagnosis and 532 standardised patient evaluations for triage. The 23 symptom checkers provided the correct diagnosis first in 34% (95% confidence interval 31% to 37%) of standardised patient evaluations, listed the correct diagnosis within the top 20 diagnoses given in 58% (55% to 62%) of standardised patient evaluations, and provided the appropriate triage advice in 57% (52% to 61%) of standardised patient evaluations. Performance on triage varied by urgency of condition, with appropriate triage advice provided in 80% (95% confidence interval 75% to 86%) of emergent cases, 55% (47% to 63%) of non-emergent cases, and 33% (26% to 40%) of self care cases (P<0.001).

Bias, confounding, and other reasons for caution

We used clinical vignettes in which the symptoms and diagnoses were typically clear, and few had comorbid conditions, resulting in a possible overestimation of the true clinical accuracy of symptom checkers. We also do not have data on the clinical performance of physicians with the same standardised patient vignettes, preventing a direct comparison between symptom checkers and physicians. When symptom checkers suggested several care sites (for example, accident and emergency department or general practice), our triage assessment was based only on the highest acuity site of care listed, and this may contribute to our finding that triage advice is risk averse.

Study funding/potential competing interests

This study was funded by the National Institutes of Health (R21 AI097759-01). The authors were independent from the funders in all aspects of the study design, analysis of data, and writing of the manuscript. All authors are affiliated with Harvard Medical School. Harvard Medical School's Family Health Guide is used as the basis for one of the symptom checkers evaluated. None of the authors have been or plan to be involved in the development or promotion of that symptom checker.

thebmj.com Read more articles about cervical screening at thebmj.com/specialties/cervical-screening

Performance of alternative strategies for primary cervical cancer screening in sub-Saharan Africa: systematic review and meta-analysis of diagnostic test accuracy studies

Joël Fokom-Domgue,[1 2] Christophe Combescure,[3] Victoire Fokom-Defo,[4] Pierre Marie Tebeu,[1] Pierre Vassilakos,[5] André Pascal Kengne,[6] Patrick Petignat[2]

[1]Department of Gynecology and Obstetrics, Faculty of Medicine and Biomedical Sciences, University of Yaoundé, Yaoundé, Cameroon
[2]Division of Gynecology, Department of Gynecology and Obstetrics, Geneva University Hospitals, 1211 Geneva 14, Switzerland
[3]Division of Clinical Epidemiology, Geneva University Hospitals, Geneva, Switzerland
[4]Division of Infectious and Chronic Diseases, Department of Internal Medicine, Yaoundé Central Hospital, Yaoundé, Cameroon
[5]Geneva Foundation for Medical Education and Research, Geneva, Switzerland
[6]South African Medical Research Council and University of Cape Town, Cape Town, South Africa
Correspondence to:
J Fokom-Domgue
fokom.domgue@gmail.com

Cite this as: *BMJ* 2015;351:h3084
doi: 10.1136/bmj.h3084

This is a summary of a paper that was published on thebmj.com as *BMJ* 2015;351:h3084

STUDY QUESTION

What is the overall accuracy (sensitivity and specificity) of cervical visual inspection with acetic acid (VIA), visual inspection with Lugol's iodine (VILI), and human papillomavirus (HPV) testing to detect cervical cancer, and how do they compare with each other in sub-Saharan Africa?

SUMMARY ANSWER

For primary cervical cancer screening in sub-Saharan Africa, VILI is a simple and affordable alternative to cytology that demonstrates higher sensitivity than VIA; although less investigated, the accuracy of HPV testing does not differ from that of VIA or VILI.

WHAT IS KNOWN AND WHAT THIS PAPER ADDS

Many accuracy studies of alternative methods for cervical cancer screening have been conducted in sub-Saharan Africa, but primary data from these studies have been conflicting. In this meta-analysis, evidence indicates that VILI performs better than VIA for primary cervical cancer screening in the region.

Selection criteria for studies

Multiple databases including Medline, Embase, and Scopus were systematically searched for studies conducted in sub-Saharan Africa and published between January 1994 and June 2014. Inclusion criteria were: alternative methods to cytology used as standalone tests for primary screening; study population not at particular risk of cervical cancer; women screened by nurses; gold standard (colposcopy and directed biopsies) performed at least in screen positive women. Two reviewers independently screened studies for eligibility, extracted data for inclusion, and evaluated the quality of eligible studies. Pooling and comparing prevalence of disease (CIN2+); positivity rates of VIA, VILI, and HPV testing; and performance measures of these screening tools used bivariate random effects models, and methods of moments.

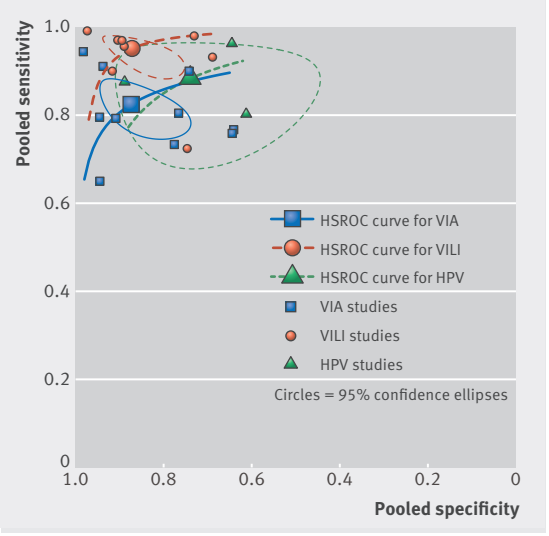
Primary outcomes

Absolute accuracy measures (sensitivity and specificity) of VIA, VILI, and HPV testing as standalone tools to detect cervical intraepithelial neoplasia grade two or worse (CIN2+).

Main results

Fifteen studies of moderate quality were identified (61 381 women for VIA, 46 435 for VILI, and 11 322 for HPV testing). Pooled sensitivity was higher for VILI (95.1%; 95% confidence interval 90.1% to 97.7%) than for VIA (82.4%; 76.3% to 87.3%) in studies where the reference test (colposcopy and directed biopsies) was performed in all women (P<0.001). Pooled specificity of VILI (87.2%; 78.1% to 92.8%) and VIA (87.4%; 77.1% to 93.4%) were similar (P=0.85). Pooled sensitivity and specificity of HPV testing were similar to VIA (both P≥0.23) and to VILI (both P≥0.16). Accuracy of VIA and VILI increased with sample size and time period. The figure shows hierarchical summary receiver operating characteristic (HSROC) regression curves, depicting fitted sensitivity as a function of specificity for VIA, VILI, and HPV testing to detect CIN2+.

Absolute accuracy of VIA, VILI, and HPV testing



Bias, confounding, and other reasons for caution

Most included studies were designed on demonstration projects and did not necessarily reflect the challenges of "real life" cervical cancer screening in Africa. VILI has never been implemented as a standalone test for primary screening, but has always been evaluated following VIA. Despite all precautions to ensure independence between the two tests in selected studies, some level of contamination cannot be eliminated. None of the selected studies used histology of a random cervical biopsy as a reference test, although the colposcopy based gold standard could lead to biased estimates of tests' performance.

Study funding/potential competing interests

Statistical analyses were funded by the International Solidarity of Geneva (Switzerland). No competing interests declared.

thebmj.com Read more articles about infectious diseases at thebmj.com/specialties/infectious-diseases

Evidence based community mobilisation for dengue prevention in Nicaragua and Mexico (*Camino Verde*, the Green Way): cluster randomised controlled trial

Neil Andersson,[1 2] Elizabeth Nava-Aguilera,[1] Jorge Arosteguí,[3] Arcadio Morales-Perez,[1] Harold Suazo-Laguna,[3] José Legorreta-Soberanis,[1] Carlos Hernandez-Alvarez,[3] Ildefonso Fernandez-Salas,[4] Sergio Paredes-Solís,[1] Angel Balmaseda,[5] Antonio Juan Cortés-Guzmán,[6] René Serrano de los Santos,[1] Josefina Coloma,[7] Robert J Ledogar,[8] Eva Harris[7]

[1]Centro de Investigación de Enfermedades Tropicales (CIET), Universidad Autónoma de Guerrero, Acapulco, Mexico
[2]Department of Family Medicine, McGill University, 5858 Côte-des-Neiges, Montreal, Canada
[3]CIET, Managua, Nicaragua
[4]Facultad de Ciencias Biológicas, Universidad Autónoma de Nuevo León, Monterrey, Mexico
[5]Laboratorio Nacional de Virología, Centro Nacional de Diagnóstico y Referencia, Ministerio de Salud, Managua, Nicaragua
[6]Departamento de Prevención y Control de Enfermedades Transmisibles por Vector, Servicios Estatales de Salud Guerrero, Av Rufo Figueroa 6, Colonia Burócratas, Chilpancingo, Mexico.
[7]Division of Infectious Diseases and Vaccinology, School of Public Health, University of California, Berkeley, Berkeley, CA, USA
[8]CIET International, New York, NY, USA

Correspondence to: N Andersson
andersson@ciet.org

Cite this as: *BMJ* 2015;351:h3267
doi: 10.1136/bmj.h3267

This is a summary of a paper that was published on thebmj.com as *BMJ* 2015;351:h3267

STUDY QUESTION

Does evidence based community engagement add effectiveness to conventional strategies for dengue control?

SUMMARY ANSWER

Evidence based community mobilisation, with each community choosing and implementing its own mix of dengue prevention actions based on local vector reservoirs and community resources, can add effectiveness to dengue vector control.

WHAT IS KNOWN AND WHAT THIS PAPER ADDS

Pesticide dependent approaches have not yet curbed the worldwide spread of dengue. This trial shows serological evidence of reduced infection with dengue virus as a result of community engagement in efforts to control mosquitoes.

Design

A parallel group, open label, cluster randomised controlled trial used central computerised randomisation after the baseline study to allocate half the clusters to intervention, stratified by country, recent dengue virus infection, and vector density.

Participants and setting

Participants were residents of a random sample of 60 clusters in the Nicaraguan capital city Managua and 90 urban and rural clusters in three coastal regions in Guerrero State in the south of Mexico. The 75 intervention and 75 control clusters included 85 182 residents in 18 838 households.

Primary outcomes

Primary per protocol outcomes were serological evidence of dengue virus infection (in children aged 3-9) in paired samples of saliva collected before and after the dengue season, self reported dengue cases, and conventional entomological indices after 18 months of intervention. Measured indices included house index (households with larvae or pupae/households examined), container index (containers with larvae or pupae/containers examined), Breteau index (containers with larvae or pupae/households examined) and pupae per person (pupae found/number of residents).

Main results and the role of chance

The figure shows the results of the main analysis, with relative risk reductions and numbers needed to treat.

Harms

Per protocol secondary analysis showed no serological evidence of a protective effect of temephos.

Bias, confounding, and other reasons for caution

We excluded 17 affluent neighbourhoods from the Managua sample before randomisation.

Generalisability to other populations

Managua and Guerrero State in Mexico cover a range of conditions in Latin America. Each site implementing the intervention in its own way has advantages of local customisation and strong community engagement. One limit to generalisability could be that the trial was implemented in both countries by an academic non-governmental organisation with three decades of experience of community engagement.

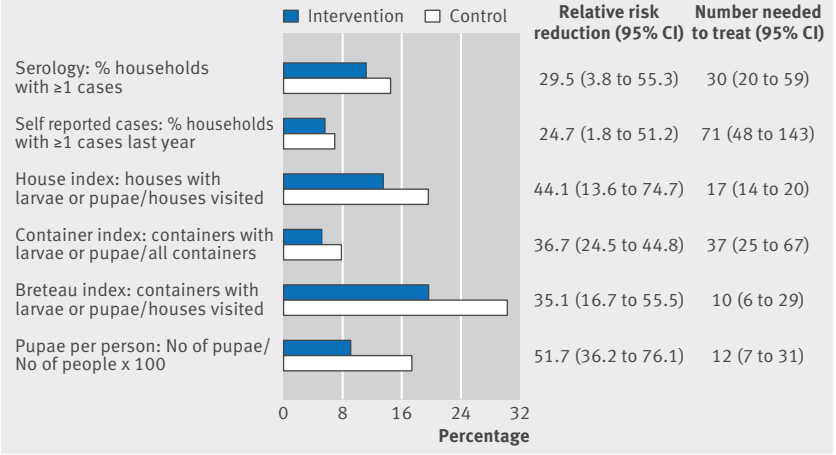
Study funding/potential competing interests

The UBS Optimus Foundation funded the study. There were no competing interests.

Trial registration number

ISRCTN27581154

Primary outcomes and confidence intervals for each group with cluster as unit of analysis



Specific SSRIs and birth defects: bayesian analysis to interpret new data in the context of previous reports

Jennita Reefhuis,[1] Owen Devine,[1] Jan M Friedman,[2] Carol Louik,[3] Margaret A Honein,[1] on behalf of the National Birth Defects Prevention Study

[1]National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, GA, USA
[2]Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada
[3]Slone Epidemiology Center at Boston University, Boston, MA, USA
Correspondence to: J Reefhuis
NZR5@cdc.gov

Cite this as: *BMJ* 2015;350:h3190
doi: 10.1136/bmj.h3190

This is a summary of a paper that was published on thebmj.com as *BMJ* 2015;350:h3190

thebmj.com

- For more on SSRIs and pregnancy go to bit.ly/1M6Kg4D
- Watch the authors talk about their findings at thebmj.com/content/350/bmj.h3190

STUDY QUESTION

Which of the previously reported associations between periconceptional use of selective serotonin reuptake inhibitors (SSRIs) and specific birth defects can be confirmed using an expanded dataset from the National Birth Defects Prevention Study (NBDPS)?

SUMMARY ANSWER

These data provide reassuring evidence for some SSRIs but suggest that some birth defects occur 2-3.5 times more often among the infants of women who were treated with paroxetine or fluoxetine early in pregnancy.

WHAT IS KNOWN AND WHAT THIS PAPER ADDS

SSRIs are increasingly used by women of reproductive age and during pregnancy, but inconsistent reports have limited opportunities for clinicians to carefully evaluate the risk compared with benefit of specific SSRIs during pregnancy. Among 21 previously reported associations between specific SSRIs and birth defects, only seven were confirmed (five for paroxetine and two for fluoxetine) and a further two had marginal associations in this analysis (one for fluoxetine and one for citalopram).

Participants and setting

The analysis included 17 952 mothers of infants with birth defects and 9857 mothers of infants without birth defects, identified through birth certificates or maternity hospitals in 10 centres participating in the NBDPS in the United States, with estimated dates of delivery between 1997 and 2009.

Design, size, and duration

We used bayesian analysis combining results from independent published analyses with data from NBDPS, a multicentre population based case-control study of birth defects.

Among 21 previously reported associations between specific SSRIs and birth defects, findings with elevated posterior odds ratios and 95% credible intervals that exclude the null value for association between specific SSRIs and birth defects, National Birth Defects Prevention Study (NBDPS), 1997-2009

Specific SSRI	Birth defect outcome	Posterior odds ratio (95% CrI)
Fluoxetine	Craniosynostosis	1.9 (1.1 to 3.0)*
Fluoxetine	Right ventricular outflow tract obstruction defect	2.0 (1.4 to 3.1)
Paroxetine	Right ventricular outflow tract obstruction defect	2.4 (1.4 to 3.9)
Paroxetine	Anencephaly	3.2 (1.6 to 6.2)
Paroxetine	Atrial septal defects	1.8 (1.1 to 3.0)
Paroxetine	Gastroschisis	2.5 (1.2 to 4.8)*
Paroxetine	Omphalocele	3.5 (1.3 to 8.0)*

*One previous literature report, which was based on a subset of NBDPS data.

Primary outcome(s), risks, exposures

The analyses included 14 birth defects categories that had associations with SSRIs reported in the literature. NBDPS data included self reported maternal use of citalopram, escitalopram, fluoxetine, paroxetine, or sertraline in the month before through the third month of pregnancy. We considered women unexposed if they did not report any antidepressants in the month before through the third month of pregnancy and did not report any depression, anxiety, bipolar disorder, or obsessive compulsive disorder.

Main results and the role of chance

Among the 21 specific SSRI-birth defects associations assessed, high posterior odds ratios with 95% credible intervals excluding the null value were observed for seven. Sertraline was the most common SSRI used in the NBDPS population, but none of the five previously reported birth defect associations for sertraline were confirmed in this analysis. We made 21 comparisons between specific SSRIs and birth defects using five different models, and some positive findings could be due to chance. Although our analysis strongly supports the validity of the associations that were observed, the increase in the absolute risks, if the associations are causal, is small. For example, if these associations are causal, the absolute risks in the children of women who are treated with paroxetine early in pregnancy would increase for right ventricular outflow tract obstruction cardiac defects from 10 per 10 000 to 24 per 10 000.

Bias, confounding, and other reasons for caution

The findings observed might be due to maternal use of SSRI, underlying maternal disease, or other unmeasured factors. A priori selected confounders adjusted for were maternal race/ethnicity, maternal education, obesity, and smoking. The three associations seen only in an earlier analysis of a subset of NBDPS data should be corroborated in an independent dataset.

Generalisability to other populations

By using findings from earlier published studies from different settings as inputs for this analysis, we further increased the generalisability of this multisite, population based study, potentially to other settings with similar healthcare provision.

Study funding/potential competing interests

The US Centers for Disease Control and Prevention funded the data collection. We have no competing interests.