BMJ Open

Diagnostic accuracy of self-administered urine glucose test strips as a diabetes screening tool in a low-resource setting in Cambodia

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019924
Article Type:	Research
Date Submitted by the Author:	09-Oct-2017
Complete List of Authors:	Storey, Helen; PATH, Diagnostics Kahn, Ansley; PATH, Noncommunicable Diseases; van Pelt, Maurits; MoPoTsyo Bun, Socheath; MoPoTsyo Daily, Frances Neogi, Tina; PATH Thompson, Matthew; University of Washington, Department of Family Medicine McGuire, Helen; PATH, Noncommunicable Diseases Weigl, Bernhard; PATH
Primary Subject Heading:	Diabetes and endocrinology
Secondary Subject Heading:	Diagnostics, Global health
Keywords:	Low-resource settings, Diabetes, Diagnostics, Urine glucose test strip, Screening

Diagnostic accuracy of self-administered urine glucose test strips as a diabetes screening tool in a low-resource setting in Cambodia

Helen L. Storey; PATH, Seattle WA 98109, USA; hstorey@path.org

Maurits H. van Pelt; MoPoTsyo, Phnom Penh, Cambodia; mopotsyo@gmail.com

Socheath Bun; MoPoTsyo, Phnom Penh, Cambodia; mopo.medtrainer@gmail.com

Frances Daily; Phnom Penh, Cambodia; fdaily@bigpond.com

Tina Neogi[1]; PATH, Seattle WA 98109, USA; tinaneogi@gmail.com

Matthew Thompson; Department of Family Medicine, University of Washington, Seattle WA 98195, USA; mjt@uw.edu

Helen McGuire; PATH, Washington DC 20001, USA; hmcguire@path.org

Bernhard H Weigl[2]; PATH, Seattle WA 98109, USA; bweigl@intven.com

* Corresponding author: Helen McGuire, PATH, PO Box 900922 Seattle WA 98109, hmcguire@path.org, 202-540-4360

[1] Present Address: UW Neighborhood Ravenna Clinic, Seattle WA 98105

[2] Present Address: Intellectual Ventures/Global Good, Bellevue WA 98005, USA

Abstract (word count: 248)

Objective: Screening for diabetes in low resource countries is a growing challenge, necessitating tests that are resource and context appropriate. The aim of this study was to determine the

diagnostic accuracy of a self-administered urine glucose test strip compared to alternative diabetes screening tools in a low resource setting of Cambodia.

Design: Prospective cross-sectional study

Setting: Members of the Borey Santhepheap community in Cambodia (Phnom Penh Municipality, District Dangkao, Commune Chom Chao).

Participants: All households on randomly selected streets were invited to participate, and adults at least 18 years of age living in the study area were eligible for inclusion.

Outcomes: The accuracy of self-administered UGTS positivity, HbA1c >6.5%, and cFBG ≥126 mg/dL were assessed against a composite reference standard of capillary FBG ≥200 mg/dL or venous blood glucose 2 hours after OGTT ≥200 mg/dL.

Results: Of the 1289 participants, 234 (18%) had diabetes based on either cFBG (74, 32%) or the OGTT (160, 68%). The UGTS was 14% sensitive and 99% specific, and failed to identify 201 individuals with diabetes, while falsely identifying 7 without diabetes. Those missed by the UGTS had lower venous FBG, lower 2-hour OGTT, and lower HbA1c compared with those correctly diagnosed.

Conclusions: Low cost, easy to use diabetes tools are essential for low-resource communities with minimal infrastructure. While the UGTS may identify persons with diabetes that might otherwise go undiagnosed in these settings, its poor sensitivity cannot be ignored. The massive burden of diabetes in low-resource settings demands improvements in test technologies.

Keywords: Diabetes, Low-resource settings, Diagnostics, Urine glucose test strip, Screening,

Article Summary (word count: 2261)

Strengths and limitations of the study

- This is one of the first studies to determine the prevalence of diabetes and report on the screening accuracy of urine glucose test strips in Cambodia, which are commonly used as screening tests in this setting.
- We used a prospective community-based design and had a large sample size with high participation rate, though participation bias towards those able to miss a day of work to attend a clinic visit may still have been an issue.
- Use of a composite reference test and not evaluating those with cFBG> 200 mg/dL by the OGTT, could have affected our study results, though the use of OGTT allows comparison of our results to those in a number of other studies.
- The urine glucose test was self-administered and self reported, which is pragmatic and aligns with the practices at MoPoTyso and other clinical settings in Cambodia, however errors in interpreting the test result could influence accuracy.

Background

According to the International Diabetes Federation (IDF), 415 million adults are living with diabetes globally, almost half of which are undiagnosed, and this number is expected to increase to 642 million by 2040.[1] As is the case for most non-communicable diseases (NCDs), three quarters of those affected live in low- and middle-income countries. In Cambodia for example, there are an estimated 230,000 people with diabetes, who are at risk for the associated micro- and macrovascular complications of this disease, including cardiovascular disease (CVD).[1,2] Strategies to reduce CVD risk may also prevent and control diabetes, which would further reduce rates of eye, kidney, and neural damage due to diabetes complications.[3] To facilitate screening

and monitoring for diabetes in these low- and middle-income countries, a low-cost, point-of-care diagnostic test that is resource and context appropriate is needed.

In low-resource settings, urine glucose test strips have been used as diabetes screening tools because they are inexpensive, noninvasive, and easy to use.[4,5] While these tests do not require fasting and are user friendly, they can only detect glucose after it has exceeded the threshold for reabsorption by the kidneys and appears in the urine. The reported threshold varies and is affected by kidney function.[6] Although their low sensitivity makes them inadequate for use as a screening tool,[7-9] the World Health Organisation (WHO) acknowledges that they may have a place in low resource settings where other tests are not possible and the prevalence of undiagnosed diabetes may be high.[9] Currently many people are not diagnosed until severe complications develop. Although the sensitivity of the urine test delays diagnosis relative to other methods, it may provide an opportunity to reduce further advancement of complications.

MoPoTsyo, a nongovernmental organization, provides screening and care services to people with diabetes and hypertension in Cambodia through an innovative, community-based peer educator model.[10-12] MoPoTsyo uses urine glucose test strips issued in the community and self-administered by patients as the initial method of diabetes screening, which has allowed them to screen over 700,000 adults, followed by confirmation with blood glucose testing for those who have a positive urine test. The aim of this study was to determine the diagnostic accuracy of a self-administered urine glucose test strip compared to alternative diabetes screening tools in a low resource setting of Cambodia. We also explored whether individuals with diabetes who were detected by urine glucose test strips differed in health status compared to those who were missed

by this test but detected by blood glucose measurement. Greater understanding of the performance of this test by the MoPoTsyo program will help to inform its optimal use.

Methods

Study design and procedures

A prospective cross-sectional study was performed among members of the Borey Santhepheap community in Cambodia (Phnom Penh Municipality, District Dangkao, Commune Chom Chao) from November 2013 to October 2014. All households on randomly selected streets were invited to participate by a local peer educator, who described the study to all potential household members. Adults at least 18 years of age living in the study area were eligible for inclusion. Individuals were excluded if they had diabetes or hypertension or had taken medications for diabetes and/or high blood pressure in the last 30 days, had kidney disease, or had received dialysis. Informed consent was obtained from all participants.The protocol was approved by the PATH Research Ethics Committee and the National Ethics Committee for Health Research (Cambodia Institutional Review Board). Study methods and results are reported in alignment with the 2015 STARD recommendations.[13]

After enrollment, all participants were screened for diabetes using a self-administered and self-reported urine glucose test strip (Sichuan Medicines and Health Products, Chengdu, China). Participants were taught how to use the test strip and read the results with assistance of a color chart, and were given several ways to report results to their peer educator. All participants were then invited to attend the clinic following an 8-hour fast for laboratory confirmed tests for diabetes and associated co-morbid risk factors. Upon arriving at the clinic all participants

provided a urine sample, a venous blood sample, and a finger stick blood sample for capillary fasting blood glucose measurement (cFBG) (On Call Plus glucometer, Acon Laboratories, San Diego, USA). If the cFBG was less than 200 mg/dL they were asked to consume a 75g oral glucose load for the oral glucose tolerance test (OGTT). The oral glucose load was ingested within 5 minutes of starting consumption, and two hours after ingestion, further venous blood and finger stick blood samples were obtained for glucose measurements. During the visit, a health history was completed based on the WHO STEPS surveillance questionnaire [14] and blood pressure measured by trained clinical staff using an ectronic device (Omron Corporation, Tokyo, Japan). All devices used in the study were owned and used previously by MoPoTsyo within the guidelines of the Cambodian Ministry of Health; none of the devices were investigational. Additional laboratory tests performed included HbA1c (DCA Vantage Analyzer, Siemens AG, Germany), serum creatinine, glucose, total cholesterol, high-density lipoprotein cholesterol, and triglycerides (Humalyzer 3000 Chemistry Analyzer, Human Diagnostics, Germany), spot urine creatinine, protein, and albumin tests (Combilyzer dipstick reader, Human Diagnostics, Germany).

A sample size of 1315 participants was calculated for a desired precision range of 10% and an estimated sensitivity and specificity of the urine glucose test strip of 21% and 90%, respectively, which is also sufficient for analysis of HbA1c, OGTT, and FBG as the test strip has the lowest performance. The sample size for the study was calculated based on Buderer's formula [15], accounting for a 3% drop-out rate and a 5% national prevalence of diabetes [16].

Data Analysis

The index tests of interest were a positive self-administered urine glucose test strip, HbA1c >6.5%, and cFBG ≥126 mg/dL. Diagnostic accuracy was assessed against a composite reference standard, which was cFBG ≥200 mg/dL, or venous blood glucose, 2 hours after OGTT ≥200 mg/dL.[17,18] If the participant's cFBG was >200 mg/dL, the patient was considered to have diabetes and an OGTT was not performed. Other measures were defined as follows: Overweight (BMI ≥25 or waist circumference >90cm for men or >80cm for women[19]), elevated blood pressure (systolic pressure ≥140mmHg or diastolic pressure ≥90mmHg), albuminuria (≥20 mg/L), and elevated albumin/creatinine ratio (≥30mg/g). We calculated sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+), negative LR (LR-),with 95% confidence intervals (CI).

Subgroup analyses were used to explore the performance of the urine glucose test strip in participants at increased risk for diabetes mellitus (DM), including age (>=50 years), BMI (>=25), gender, and waist circumference (>90cm for men or >80cm for women). Prevalence of diabetes by subgroup was compared by chi-squared test. We also explored whether the individuals correctly classified by the urine glucose test strip had better or worse controlled diabetes than those misclassified by the test, as defined by various clinical and laboratory measures. Continuous values were compared using Student's t-test and dichotomous values were compared using the chi-squared test. Data were analyzed using Stata/SE 13.1 (StataCorp LP, Texas, USA).

Results

Of 1328 eligible study subjects, 1316 participated in the study and 1289 were included in the analysis (Figure 1). Participants were excluded from the analysis if they did not complete the OGTT due to vomiting or other reasons (16),were not fasting prior to the clinic visit (5), or reported taking medication for diabetes that day (6). Of the analyzed participants, 75% (972/1289) were female, mean age was 51 years, 31% had high BMI, and 13% had elevated blood pressure, although only 8% were taking antihypertensive medications. Characteristics of the participants included in the analysis are presented in Table 1.

A total of 234 individuals had diabetes based on the composite reference standard of either cFBG(74, 32%) or the OGTT (160, 68%), corresponding to a prevalence of 18%. Of the index tests evaluated, the urine glucose test strip had lower sensitivity (14.1% sensitive), than cFBG (73.9%), and HbA1c (75.2% sensitive). All three tests offered high specificity (99.3%, 96.8% and 98.5% respectively) (Table 2).The urine glucose test strip failed to identify 201 individuals with diabetes (false negatives) and falsely identified seven participants without diabetes (false positives). The 201 patients with diabetes who were not identified by the urine test had significantly lower venous FBG, lower 2 hr OGTT, and lower HbA1c compared to those correctly diagnosed, but were similar in other characteristics (Table 3). The seven false positive individuals had higher HbA1c, higher systolic BP, and higher proportion receiving treatment for hypertension than those with true negative results (Table 3).

The prevalence of diabetes (diagnosed by the composite reference standard) was significantly higher in participants who were 50 years of age or older compared to those under 50 years (24% vs. 9.6%, $p<0.001$); those with high BMI compared to those with normal BMI (22% vs. 17%,

p=0.03); and those with greater waist circumference compared to those with normal waist (24% vs. 13%, p<0.001), but was the same in males and females (Table 4). The diagnostic accuracy of the urine glucose test strip was similar among subgroups of patients with various cofactors, with overlapping confidence intervals (Table 4).

Discussion

Urine glucose test strips had much lower sensitivity than either cFBG or HbA1c, but all three tests offered high specificity. Patients who tested positive with the urine glucose test who were confirmed to have diabetes by the reference standard (true positives) had higher FBG, higher OGTT and higher HbA1c levels compared to the false negative group (urine test negative in patients with diabetes), suggesting that the urine glucose test may identify individuals with poor glycemic control. This suggests a subset of diabetes patients is being identified that is potentially at higher risk of advancing complications or comorbidities, and who may benefit the most from further care [20]. In addition, testing for urine glucose was highly specific (99%), with positive LRs in the 20s, indicating that when positive, this test is highly indicative of diabetes.

The prevalence of diabetes in the MoPoTsyo population in Cambodia was 18%. This is much higher than the national prevalence for Cambodia, which is reported at 3.0%.[1] This may be due to the high proportion of individuals over 50 years of age in our study population, which could be explained by a participation bias towards those who were able to miss a day of work to attend a clinic visit. Additionally, our study took place in a rapidly changing urban population, which had a 2.4 times higher diabetes prevalence in the STEP survey, country report from 2010.[21]

A wide range of sensitivities for the urine glucose test strip has been reported, and its use remains controversial. A review in 2000 found six adequately designed studies that reported performance of urine test strips for glucose.[8] Among these, sensitivities in two reports of fasting patients were 16% and 35%; two using random samples found sensitivities of 18% and 64%; and three using postprandial and post-load measurements reported sensitivities between 39% and 48%. This review concluded that blood glucose measurements were preferred over urinary glucose or HbA1c, and particularly, postprandial over fasting measures. Another review found five studies reporting a range of sensitivity from 18% to 74% for urine glucose test strips.[7] The review concluded that urine glucose test strips are not sufficient for screening for diabetes.

This is one of the first studies to determine the prevalence of diabetes in Cambodia, and report on the screening accuracy of urine glucose test strips which are commonly used as screening tests in this setting. We used a prospective community-based design and had a large sample size with high participation rate. The study had several limitations. Firstly, we used a composite reference test and those with cFBG> 200 mg/dL were not evaluated by the OGTT. While OGTT is considered the gold standard reference test for assessing diagnostic accuracy, there has been some question of its performance. Two studies in China, each on more than 200 participants, found that the reproducibility of the OGTT was 56% [22] and 66% [23]. Though our choice of the reference standards, particularly OGTT, could have affected our study results, its use allows comparison of our results to those in a number of other studies. Second, the urine glucose test was self-administered and self reported. While this was pragmatic, and aligns with the practices at MoPoTyso and other clinical settings in Cambodia, errors in interpreting the test result could

influence accuracy. We were not able to repeat this test when patients attended their clinic visit as they were fasting at the clinic visit, and thus their urine would not have been the random non-fasting urine test obtained at home. Third, we were not able to obtain hemoglobin levels (or test for hemoglobin variants) as these tests are not available in this setting, and hence cannot assess the impact of anemia or hemoglobinopathy on test performance. Fourth, glucose test strip accuracy may be subject to effects of heat and humidity, we were not able to explore their possible impact on our results.

For clinicians working in settings similar to ours, the question is how useful is the urine glucose test as a screening or diagnostic test, and is it "better than nothing"? The low sensitivity certainly reduces the value of this test as a screening tool, but the high specificity means that positive tests can be used to rule in patients with diabetes, suggesting that urine glucose may have some diagnostic value in this setting. The false positive rate was extremely low, and only 7 patients without disease were identified as positive by urine glucose test strip. From a population perspective, the value of a low cost, poorly sensitive yet highly specific test for diabetes is unclear in terms of balancing the opportunity to identify a subset of patients with less well controlled diabetes who would not have been identified otherwise, with the downside of a high false negative rate.[24]

Not surprisingly, usability parameters and cost make urine glucose test strips a highly desirable test in this and other low-resource settings.[9] Product attributes such as low complexity and infrastructure requirements, short time to results, and low participant burden greatly contribute to the acceptability and desirability of the screening tool. The large patient burden and the frequent

inability to comply with fasting requirements reduce the feasibility of using OGTT or FBG tests. While HbA1c testing does not require fasting, current tests are too expensive for use in most low-income countries. The role of a poorly sensitive test like urine glucose in resource poor settings such as Cambodia is debatable, on the one hand the test will identify some patients previously undiagnosed, and assuming treatment can be initiated, reduce severity of complications from this disease. On the other hand, the test will miss the majority of patients with diabetes, thus risking a false reassurance, further postponement of diagnosis, and risking patient's respect for the health care system.

There may be strategies to improve the performance (particularly sensitivity) of the urine glucose test strip. First, using presence of risk factors such as high waist circumference or BMI, may increase the pretest probability of diabetes and lead to improved performance. Second, using random, postprandial, or glucose-loaded measurements may be superior than fasting because the renal threshold for glucose is more often reached in non-fasting states.[8] Third, improving the limit of detection may be possible by modifications in the test strip itself, or improvement in the way it is read either manually (with trained users) or automatically (with electronic reading devices). Finally, increasing screening frequency may be feasible in low resource settings, if the urine glucose test strip truly does identify a smaller but more advanced fraction of diabetes patients.

Conclusion

Low cost, easy to use diabetes screening, diagnosis, and monitoring tools are essential for low-resource communities with minimal infrastructure. While the urine glucose test strip has some

value as a screening test in these settings, its performance is far from optimal. Progress is urgently needed to improve the performance, availability, and access of essential testing technologies for diabetes.

List of abbreviations

urine glucose test strip (UGTS)

International Diabetes Federation (IDF)

non-communicable diseases (NCDs)

cardiovascular disease (CVD)

World Health Organisation (WHO)

capillary fasting blood glucose measurement (cFBG)

oral glucose tolerance test (OGTT)

positive predictive value (PPV)

negative predictive value (NPV)

positive likelihood ratio (LR+)

negative likelihood ratio (LR-)

confidence intervals (CI)

diabetes mellitus (DM)

Declarations

Ethical approval and consent to participate

The protocol was approved by the PATH Research Ethics Committee and the National Ethics Committee for Health Research (Cambodia Institutional Review Board). Informed consent was obtained from all participants.

Consent for publication

Not applicable.

Availability of data and material

The datasets used during the current study are available from the corresponding author on reasonable request.

Competing Interests

The authors declare that they have no competing interests.

Funding

This work was supported by a grant from Medtronic Foundation, and received additional support from PATH and the University of Washington Department of Family Medicine. The funding source had no involvement in study design, data collection, data analysis, data interpretation, writing of the manuscript, or the decision to publish the results.

Authors contributions

MHP, SB, TN, HM and BW designed the study; MHP, SB, TN, and BW implemented the study; HLS, MT, HM, and BW analysed and interpreted the data; HLS, MHP, FD, MT, HM, and BW contributed to writing. All authors read and approved the final manuscript.

Acknowledgements

We would like to acknowledge the Borey Santhepheap community in Cambodia (Phnom Penh Municipality, District Dangkao, Commune Chom Chao) for participating in this study. We also

acknowledge the input of Dr Annette Fitzpatrick and Dr Jim LoGerfo from the University of Washington.

Authors' information

Not applicable.

Tables

Table 1. Characteristics of included participants.

	Mean (SD) or % n=1289
Age, years	51.4 (14.9)
Female (%)	75.4
BMI[1]	23.2 (4.1)
High BMI (%)	30.5
Waist circumference above cutoff [2] (%)	46.1
Systolic blood pressure, mmHg	123.5 (20.6)
Diastolic blood pressure, mmHg	80.8 (12.1)
Elevated blood pressure (%)	12.9
Take treatment for high blood pressure (%)	8.2

[1] n=1288

[2] >90cm for men, >80cm for women. [19]

Table 2. Diagnostic accuracy of urine glucose test strip, capillary fasting glucose, and HbA1c determined by comparison with the composite reference standard (n=1289)[1].

	Urine glucose test strip positive	**cFBG ≥126 mg/dL**	**HbA1c >6.5%**
True positive (n)	33	173	176
False positive (n)	7	34	16
False negative (n)	201	61	58
True negative (n)	1048	1021	1039
True diabetes prevalence[1] (95%CI)	18%, 234/1289 (16, 20.4)		
Sensitivity (95% CI)	14.1 (9.90, 19.2)	73.9 (67.8, 79.4)	75.2 (69.2, 80.6)
Specificity (95% CI)	99.3 (98.6, 99.7)	96.8 (95.5, 97.8)	98.5 (97.5, 99.1)
Positive PV (95% CI)	82.5 (67.2, 92.7)	83.6 (77.8, 88.3)	91.7 (86.8, 95.2)
Negative PV (95% CI)	83.9 (81.7, 85.9)	94.4 (92.8, 95.7)	94.7 (93.2, 96.0)
Positive LR (95% CI)	21.3 (9.50, 47.5)	22.9 (16.3, 32.2)	49.6 (30.3, 81.1)
Negative LR (95% CI)	0.90 (0.80, 0.90)	0.30 (0.20, 0.30)	0.30 (0.20, 0.30)

[1] Excludes individuals taking diabetes treatment that day (n=6), did not fast before OGTT as instructed (n=5), or did not complete the OGTT (n=16). 74 patients with cFBG>=200 were not tested by OGTT.

[2] Composite reference standard: OGTT ≥200 mg/dL or cFBG ≥200 mg/dL.

Table 3. Diagnostic accuracy of the urine glucose test strip by patient characteristics.

Patient characteristic	Diabetic[1]		Non-diabetic[1]	
	True Positive n=33 Mean (SD) or %	False Negative n=201 Mean (SD) or %	False Positive n=7 Mean (SD) or %	True Negative n=1048 Mean (SD) or %
Age	57 (9.3)	58 (10.5)	56 (11.9)	50 (15.5)
Female (%)	81.8	74.6	85.7	75.3
Venous fasting blood glucose	**207 (75.3)**	**166 (73.2)**	95 (16.9)	90 (13.1)
Venous blood glucose 2 hrs after OGTT	**310 (60.8)**	**275 (62.2)**	115 (43.2)	120 (31.0)
Change in venous blood glucose during OGTT	160 (50.8)	146 (49.8)	20 (47.7)	30 (30.0)
HbA1c	**10 (2.3)**	**8 (2.4)**	**6 (0.7)**	**5 (0.5)**
BMI	24 (3.9)	24 (3.9)	26 (3.2)	23 (4.1)
High BMI (%)	33.3	36.8	57.1	29.0
Waist circumference above cutoff (%)	60.6	61.7	71.4	42.8
Systolic blood pressure	132 (24.9)	130 (20.6)	**146 (14.0)**	**122 (20.2)**
Diastolic blood pressure	85 (9.6)	84 (11.7)	87 (6.5)	80 (12.1)
Elevated blood pressure (%)	15.2	20.9	14.3	11.3
Take treatment for high blood pressure (%)	18.2	11.4	**28.6**	**7.1**
Total Cholesterol	242 (62.3)	227 (69.8)	240 (63.1)	213 (56.3)
Proteinuria (n=1116)[2] (%)	20.0	17.2	0	3.0
Albuminuria (%)	51.5	47.8	14.3	21.7
Abnormal albumin/creatinine ratio (%)	39.3	39.3	14.3	17.3

[1] Diagnosis by the composite reference standard: venous OGTT ≥200 mg/dL or cFBG ≥200 mg/dL.
[2] 4 missing values, 169 indeterminate measurements not included in analysis.
Bold = significantly different ($p \leq 0.05$) by Student's t-test or chi-squared test.

Table 4. Diagnostic accuracy of urine glucose test strip by participant cofactors (n=1289) [1].

Results	Cofactors							
	Age		BMI [1]		Gender		Waist circumference [3]	
	<50	≥50	<25	≥25	Male	Female	Normal	High
Number of participants	531	758	895	393	317	972	691	598
True positive (n)	8	25	22	11	6	27	13	20
False positive (n)	3	4	3	4	1	6	2	5
False negative (n)	43	158	127	74	51	150	77	124
True negative (n)	477	571	743	304	259	789	599	449
True diabetes prevalence [2]	**9.6% (7.2, 12.4)**	**24% (21.0, 27.4)**	**17% (14.0, 19.3)**	**22% (18.0, 26.0)**	18% (14.0, 22.7)	18% (16.0, 20.8)	**13% (11.0, 15.8)**	**24% (21.0, 27.7)**
Sensitivity (95% CI)	15.7 (7.0, 28.6)	13.7 (9.0, 19.5)	14.8 (9.5, 21.5)	12.9 (6.6, 22.0)	10.5 (4.0, 21.5)	15.3 (10.3, 21.4)	14.4 (7.9, 23.4)	13.9 (8.7, 20.6)

Specificity (95% CI)	99.4 (98.2, 99.9)	99.3 (98.2, 99.8)	99.6 (98.8, 99.9)	98.7 (96.7, 99.6)	99.6 (97.9, 100)	99.2 (98.4, 99.7)	99.7 (98.8, 100)	98.9 (97.4, 99.6)
Positive PV (95% CI)	72.7 (39, 94.0)	86.2 (68.3, 96.1)	88 (68.8, 97.5)	73.3 (44.96, 92.2)	85.7 (42.1, 99.6)	81.8 (64.5, 93.0)	86.7 (59.5, 98.3)	80 (59.3, 93.2)
Negative PV (95% CI)	91.7 (89, 94)	78.3 (75.2, 81.3)	85.4 (82.9, 87.7)	80.4 76.1, 84.3)	83.5 (78.9, 87.5)	84 (81.5, 86.3)	88.6 (86, 90.9)	78.4 (74.8, 81.7)
Positive LR (95% CI)	25.1 (6.9, 91.6)	19.6 (6.9, 55.7)	36.7 (11.1, 121)	10.0 (3.3, 30.5)	27.4 (3.4, 223)	20.2 (8.5, 48.2)	43.4 (10.0, 189)	12.6 (4.8, 33)
Negative LR (95% CI)	0.8 (0.8, 1.0)	0.9 (0.8, 0.9)	0.9 (0.8, 0.9)	0.9 (0.8, 1.0)	0.9 (0.8, 1.0)	0.85 (0.80, 0.91)	0.86 (0.79, 0.94)	0.87 (0.82, 0.93)

[1] Excluded individuals taking diabetes treatment that day (n=6), did not fast before OGTT as instructed (n=5), or did not complete the OGTT (n=16).74 patients with cFBG>=200 were not tested by OGTT; 1 patient had cFBG>=200 and also tested OGTT positive.

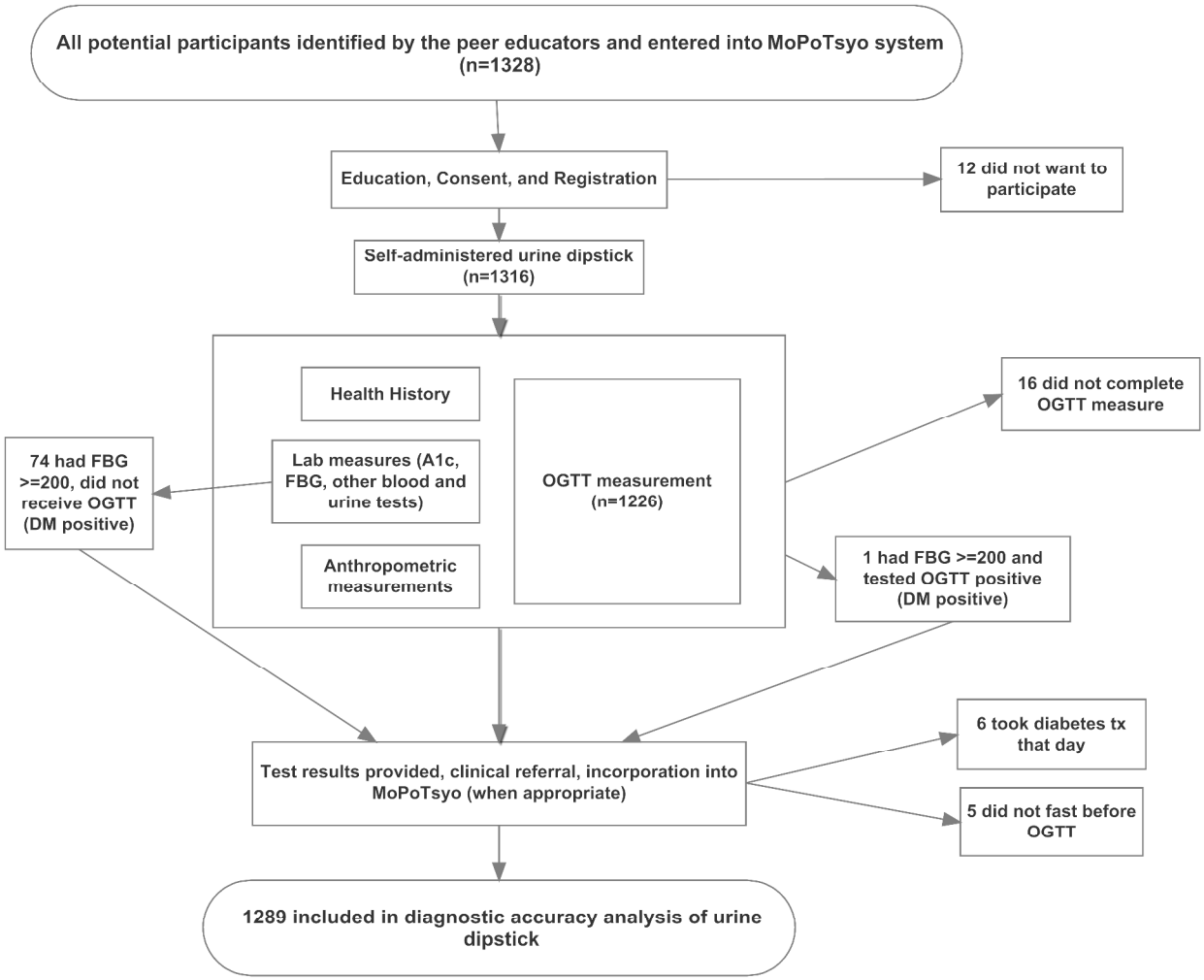
[2] True prevalence as determined by the composite reference standard. Total number of diabetes diagnoses: 234 (18% prevalence).

[3] High Waist circumference = >90cm for men, >80cm for women.[19]

Bold = significantly different ($p \leq 0.05$), chi-squared test.

Figures

Figure 1: Study flow diagram.



Reference List

1. IDF Diabetes Atlas, 7th Edition. International Diabetes Federation . 2015. International Diabetes Federation.
Ref Type: Electronic Citation

2. Forbes JM, Cooper ME: **Mechanisms of diabetic complications.** *Physiol Rev* 2013, **93(1):** 137-188.

3. Tuomilehto J, Lindstrom J, Eriksson JG, Valle TT, Hamalainen H, Ilanne-Parikka P *et al.*: **Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance.** *N Engl J Med* 2001, **344:** 1343-1350.

4. Dyerberg J, Pedersen L, Aagaard O: **Evaluation of a dipstick test for glucose in urine.** *Clin Chem* 1976, **22:** 205-210.

5. van der Sande MA, Walraven GE, Bailey R, Rowley JT, Banya WA, Nyan OA *et al.*: **Is there a role for glycosuria testing in sub-Saharan Africa?** *Trop Med Int Health* 1999, **4:** 506-513.

6. Taal MW, Chertow GM, Marsden PA, Skorecki K, Yu ASL, Brenner BM: **Laboratory Assessment of Kidney Disease: Glomerular Filtration Rate, Urinalysis, and Proteinuria.** In *Brenner & Rector's The Kidney*. Philadelphia: Elsevier; 2012.

7. Wei OY, Teece S: **Best evidence topic report. Urine dipsticks in screening for diabetes mellitus.** *Emerg Med J* 2006, **23:** 138.

8. Engelgau MM, Narayan KM, Herman WH: **Screening for type 2 diabetes.** *Diabetes Care* 2000, **23:** 1563-1580.

9. World Health Organization. Screening for Type 2 Diabetes. Report of a World Health Organization and International Diabetes Federation meeting. 2003. Geneva, Switzerland, World Health Organization.

Ref Type: Report

10. van Pelt M: **Improving access to education and care in Cambodia.** *Diabetes Voice* 2009, **54**.

11. Taniguchi D, LoGerfo J, van PM, Mielcarek B, Huster K, Haider M *et al.*: **Evaluation of a multi-faceted diabetes care program including community-based peer educators in Takeo province, Cambodia, 2007-2013.** *PLoS One* 2017, **12:** e0181582.

12. van Olmen J, Eggermont N, van Pelt M, Hen H, de Man J, Schellevis F *et al.*: **Patient-centred innovation to ensure access to diabetes care in Cambodia: the case of MoPoTsyo.** *J Pharm Policy Pract* 2016, **9:** 1.

13. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L *et al.*: **STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies.** *BMJ* 2015, **351:** h5527.

14. World Health Organization. WHO STEPS Surveillance Manual. World Health Organization . 6-13-2008. World Health Organization. 4-30-2015.

Ref Type: Electronic Citation

15. Obuchowski NA: **Sample size calculations in studies of test accuracy.** *Stat Methods Med Res* 1998, **7:** 371-392.

16. King H, Keuky L, Seng S, Khun T, Roglic G, Pinget M: **Diabetes and associated disorders in Cambodia: two epidemiological surveys.** *Lancet* 2005, **366:** 1633-1639.

17. **Standards of Medical Care in Diabetes-2016 Abridged for Primary Care Providers.** *Clin Diabetes* 2016, **34:** 3-21.

18. Bartoli E, Fra GP, Carnevale Schianca GP: **The oral glucose tolerance test (OGTT) revisited.** *Eur J Intern Med* 2011, **22:** 8-12.

19. World Health Organization. Waist circumference and waist-hip ratio: report of a WHO expert consultation, Geneva, 8-11 December 2008. 2011. Geneva, Switzerland, World Health Organization.

Ref Type: Report

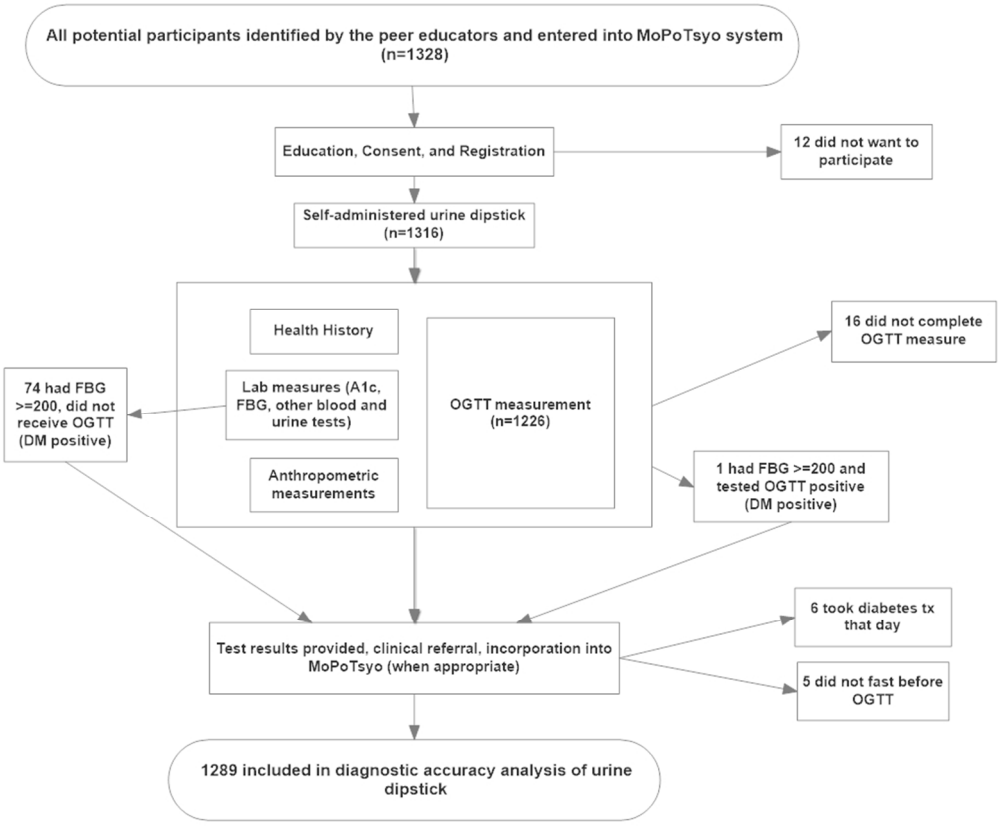
20. **The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The Diabetes Control and Complications Trial Research Group.** *N Engl J Med* 1993, **329:** 977-986.

21. University of Health Sciences Cambodia, Ministry of Health Cambodia. Prevalence of non-communicable disease risk factors in Cambodia; STEPS Survey Country Report. 2010.
Ref Type: Report

22. Liu M, Pan CY, Jin MM, Su HY, Lu JM: **[The reproducibility and clinical significance of oral glucose tolerance test for abnormal glucose metabolism].** *Zhonghua Nei Ke Za Zhi* 2007, **46:** 1007-1010.

23. Ko GT, Chan JC, Woo J, Lau E, Yeung VT, Chow CC *et al.*: **The reproducibility and usefulness of the oral glucose tolerance test in screening for diabetes and other cardiovascular risk factors.** *Ann Clin Biochem* 1998, **35 (Pt 1):** 62-67.

24. Flessa S, Zembok A: **Costing of diabetes mellitus type II in Cambodia.** *Health Econ Rev* 2014, **4:** 24.



86x71mm (300 x 300 DPI)

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	1
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	3
	4	Study objectives and hypotheses	4
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	5
Participants	**6**	Eligibility criteria	5
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	5
	8	Where and when potentially eligible participants were identified (setting, location and dates)	5
	9	Whether participants formed a consecutive, random or convenience series	5
Test methods	**10a**	Index test, in sufficient detail to allow replication	5
	10b	Reference standard, in sufficient detail to allow replication	6
	11	Rationale for choosing the reference standard (if alternatives exist)	6
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	6
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	6
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	6
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	6
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	7
	15	How indeterminate index test or reference standard results were handled	7
	16	How missing data on the index test and reference standard were handled	7
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	7
	18	Intended sample size and how it was determined	6
RESULTS			
Participants	**19**	Flow of participants, using a diagram	17
	20	Baseline demographic and clinical characteristics of participants	15
	21a	Distribution of severity of disease in those with the target condition	15
	21b	Distribution of alternative diagnoses in those without the target condition	NA
	22	Time interval and any clinical interventions between index test and reference standard	5
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	15
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	15
	25	Any adverse events from performing the index test or the reference standard	NA
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	10
	27	Implications for practice, including the intended use and clinical role of the index test	11
OTHER INFORMATION			
	28	Registration number and name of registry	NA
	29	Where the full study protocol can be accessed	NA
	30	Sources of funding and other support; role of funders	14

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.

BMJ Open

Diagnostic accuracy of self-administered urine glucose test strips as a diabetes screening tool in a low-resource setting in Cambodia

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019924.R1
Article Type:	Research
Date Submitted by the Author:	09-Jan-2018
Complete List of Authors:	Storey, Helen; PATH, Diagnostics Kahn, Ansley; PATH, Noncommunicable Diseases; van Pelt, Maurits; MoPoTsyo Bun, Socheath; MoPoTsyo Daily, Frances Neogi, Tina; PATH Thompson, Matthew; University of Washington, Department of Family Medicine McGuire, Helen; PATH, Noncommunicable Diseases Weigl, Bernhard; PATH
Primary Subject Heading:	Diabetes and endocrinology
Secondary Subject Heading:	Diagnostics, Global health
Keywords:	Low-resource settings, Diabetes, Diagnostics, Urine glucose test strip, Screening

Diagnostic accuracy of self-administered urine glucose test strips as a diabetes screening tool in a low-resource setting in Cambodia

Helen L. Storey; PATH, Seattle WA 98109, USA; hstorey@path.org

Maurits H. van Pelt; MoPoTsyo, Phnom Penh, Cambodia; mopotsyo@gmail.com

Socheath Bun; MoPoTsyo, Phnom Penh, Cambodia; mopo.medtrainer@gmail.com

Frances Daily; Phnom Penh, Cambodia; fdaily@bigpond.com

Tina Neogi[1]; PATH, Seattle WA 98109, USA; tinaneogi@gmail.com

Matthew Thompson; Department of Family Medicine, University of Washington, Seattle WA 98195, USA; mjt@uw.edu

Helen McGuire; PATH, Washington DC 20001, USA; hmcguire@path.org

Bernhard H Weigl[2]; PATH, Seattle WA 98109, USA; bweigl@intven.com

* Corresponding author: Helen McGuire, PATH, PO Box 900922 Seattle WA 98109, hmcguire@path.org, 202-540-4360

[1] Present Address: UW Neighborhood Ravenna Clinic, Seattle WA 98105

[2] Present Address: Intellectual Ventures/Global Good, Bellevue WA 98005, USA

Abstract (word count: 287)

Objective: Screening for diabetes in low resource countries is a growing challenge, necessitating tests that are resource and context appropriate. The aim of this study was to determine the

diagnostic accuracy of a self-administered urine glucose test strip compared to alternative diabetes screening tools in a low resource setting of Cambodia.

Design: Prospective cross-sectional study

Setting: Members of the Borey Santhepheap community in Cambodia (Phnom Penh Municipality, District Dangkao, Commune Chom Chao).

Participants: All households on randomly selected streets were invited to participate, and adults at least 18 years of age living in the study area were eligible for inclusion.

Outcomes: The accuracy of self-administered urine glucose test strip positivity, HbA1c >6.5%, and capillary fasting blood glucose measurement ≥126 mg/dL were assessed against a composite reference standard of capillary fasting blood glucose measurement ≥200 mg/dL or venous blood glucose 2 hours after oral glucose tolerance test ≥200 mg/dL.

Results: Of the 1289 participants, 234 (18%) had diabetes based on either capillary fasting blood glucose measurement (74, 32%) or the oral glucose tolerance test (160, 68%). The urine glucose test strip was 14% sensitive and 99% specific, and failed to identify 201 individuals with diabetes, while falsely identifying 7 without diabetes. Those missed by the urine glucose test strip had lower venous fasting blood glucose, lower venous blood glucose 2 hours after oral glucose tolerance test, and lower HbA1c compared with those correctly diagnosed.

Conclusions: Low cost, easy to use diabetes tools are essential for low-resource communities with minimal infrastructure. While the urine glucose test strip may identify persons with diabetes that might otherwise go undiagnosed in these settings, its poor sensitivity cannot be ignored. The massive burden of diabetes in low-resource settings demands improvements in test technologies.

Keywords: Diabetes, Low-resource settings, Diagnostics, Urine glucose test strip, Screening,

Article Summary (word count: 2261)

Strengths and limitations of the study

- This is one of the first studies to determine the prevalence of diabetes and report on the screening accuracy of urine glucose test strips in Cambodia, which are commonly used as screening tests in this setting.
- We used a prospective community-based design and had a large sample size with high participation rate, though participation bias towards those able to miss a day of work to attend a clinic visit may still have been an issue.
- Use of a composite reference test and not evaluating those with cFBG> 200 mg/dL by the OGTT, could have affected our study results, though the use of OGTT allows comparison of our results to those in a number of other studies.
- The urine glucose test was self-administered and self reported, which is pragmatic and aligns with the practices at MoPoTyso and other clinical settings in Cambodia, however errors in interpreting the test result could influence accuracy.

Background

According to the International Diabetes Federation (IDF), 415 million adults are living with diabetes globally, almost half of which are undiagnosed, and this number is expected to increase to 642 million by 2040.[1] As is the case for most non-communicable diseases (NCDs), three quarters of those affected live in low- and middle-income countries. In Cambodia for example, there are an estimated 230,000 people with diabetes, who are at risk for the associated micro- and macrovascular complications of this disease, including cardiovascular disease (CVD).[1,2] Strategies to reduce CVD risk may also prevent and control diabetes, which would further reduce

rates of eye, kidney, and neural damage due to diabetes complications.[3] To facilitate screening and monitoring for diabetes in these low- and middle-income countries, a low-cost, point-of-care diagnostic test that is resource and context appropriate is needed.

In low-resource settings, urine glucose test strips have been used as diabetes screening tools because they are inexpensive, noninvasive, and easy to use.[4,5] While these tests do not require fasting and are user friendly, they can only detect glucose after it has exceeded the threshold for reabsorption by the kidneys and appears in the urine. The reported threshold varies and is affected by kidney function.[6] Although their low sensitivity makes them inadequate for use as a screening tool,[7-9] the World Health Organisation (WHO) acknowledges that they may have a place in low resource settings where other tests are not possible and the prevalence of undiagnosed diabetes may be high.[9] Currently many people are not diagnosed until severe complications develop. Although the sensitivity of the urine test delays diagnosis relative to other methods, it may provide an opportunity to reduce further advancement of complications.

MoPoTsyo, a nongovernmental organization, provides screening and care services to people with diabetes and hypertension in Cambodia through an innovative, community-based peer educator model.[10-12] MoPoTsyo uses urine glucose test strips issued in the community and self-administered by patients as the initial method of diabetes screening, which has allowed them to screen over 700,000 adults, followed by confirmation with blood glucose testing for those who have a positive urine test. The aim of this study was to determine the diagnostic accuracy of a self-administered urine glucose test strip compared to alternative diabetes screening tools in a low resource setting of Cambodia. We also explored whether individuals with diabetes who were

detected by urine glucose test strips differed in health status compared to those who were missed by this test but detected by blood glucose measurement. Greater understanding of the performance of this test by the MoPoTsyo program will help to inform its optimal use.

Methods

Study design and procedures

A prospective cross-sectional study was performed among members of the Borey Santhepheap community in Cambodia (Phnom Penh Municipality, District Dangkao, Commune Chom Chao) from November 2013 to October 2014. All households on randomly selected streets were invited to participate by a local peer educator, who described the study to all potential household members. Adults at least 18 years of age living in the study area were eligible for inclusion. Individuals were excluded if they had diabetes or hypertension or had taken medications for diabetes and/or high blood pressure in the last 30 days, had kidney disease, or had received dialysis. Written informed consent was obtained from all participants.The protocol was approved by the PATH Research Ethics Committee and the National Ethics Committee for Health Research (Cambodia Institutional Review Board). Study methods and results are reported in alignment with the 2015 STARD recommendations.[13]

After enrollment, all participants were screened for diabetes using a self-administered and self-reported urine glucose test strip (Sichuan Medicines and Health Products, Chengdu, China). Participants were taught how to use the test strip and read the results with assistance of a color chart, and were given several ways to report results to their peer educator. All participants were then invited to attend the clinic following an 8-hour fast for laboratory confirmed tests for

diabetes and associated co-morbid risk factors. Upon arriving at the clinic all participants provided a urine sample, a venous blood sample, and a finger stick blood sample for capillary fasting blood glucose measurement (cFBG) (On Call Plus glucometer, Acon Laboratories, San Diego, USA, https://www.aconlabs.com/us/glucose/on-call/plus-bgms/). If the cFBG was less than 200 mg/dL they were asked to consume a 75g oral glucose load for the oral glucose tolerance test (OGTT). The oral glucose load was ingested within 5 minutes of starting consumption, and two hours after ingestion, further venous blood and finger stick blood samples were obtained for glucose measurements. During the visit, a health history was completed based on the WHO STEPS surveillance questionnaire [14] and blood pressure measured by trained clinical staff using an ectronic device (Omron Corporation, Tokyo, Japan). All devices used in the study were owned and used previously by MoPoTsyo within the guidelines of the Cambodian Ministry of Health; none of the devices were investigational. Additional laboratory tests performed included HbA1c (DCA Vantage Analyzer, Siemens AG, Germany), serum creatinine, glucose, total cholesterol, high-density lipoprotein cholesterol, and triglycerides (Humalyzer 3000 Chemistry Analyzer, Human Diagnostics, Germany), spot urine creatinine, protein, and albumin tests (Combilyzer dipstick reader, Human Diagnostics, Germany).

A sample size of 1315 participants was calculated for a desired precision range of 10% and an estimated sensitivity and specificity of the urine glucose test strip of 21% and 90%, respectively, which is also sufficient for analysis of HbA1c, OGTT, and FBG as the test strip has the lowest performance. The sample size for the study was calculated based on Buderer's formula [15], accounting for a 3% drop-out rate and a 5% national prevalence of diabetes [16].

Data Analysis

The index tests of interest were a positive self-administered urine glucose test strip, HbA1c >6.5%, and cFBG ≥126 mg/dL. Diagnostic accuracy was assessed against a composite reference standard, which was cFBG ≥200 mg/dL, or venous blood glucose, 2 hours after OGTT ≥200 mg/dL.[17,18] If the participant's cFBG was >200 mg/dL, the patient was considered to have diabetes and an OGTT was not performed. Other measures were defined as follows: Overweight (BMI ≥25 or waist circumference >90cm for men or >80cm for women[19]), elevated blood pressure (systolic pressure ≥140mmHg or diastolic pressure ≥90mmHg), albuminuria (≥20 mg/L), and elevated albumin/creatinine ratio (≥30mg/g). We calculated sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+), negative LR (LR-),with 95% confidence intervals (CI).

Subgroup analyses were not prespecified, and therefore used to explore the performance of the urine glucose test strip in participants at increased risk for diabetes mellitus (DM), including age (>=50 years), BMI (>=25), gender, and waist circumference (>90cm for men or >80cm for women). Logistic regression analyses were also used to determine if the diagnostic accuracy of the index test was impacted by these clinical features. Prevalence of diabetes by subgroup was compared by chi-squared test. We also explored whether the individuals correctly classified by the urine glucose test strip had better or worse controlled diabetes than those misclassified by the test, as defined by various clinical and laboratory measures. Mean values of continuous variables were compared using Student's t-test while proportions of dichotomous values were compared using the chi-squared test. Data were analyzed using Stata/SE 13.1 (StataCorp LP, Texas, USA).

Results

Of 1328 eligible study subjects, 1316 participated in the study and 1289 were included in the analysis (Figure 1). Participants were excluded from the analysis if they did not complete the OGTT due to vomiting or other reasons (16), were not fasting prior to the clinic visit (5), or reported taking medication for diabetes that day (6). Of the analyzed participants, 75% (972/1289) were female, mean age was 51 years, 31% had high BMI, and 13% had elevated blood pressure, although only 8% were taking antihypertensive medications. Characteristics of the participants included in the analysis are presented in Table 1.

A total of 234 individuals had diabetes based on the composite reference standard of either cFBG (70, 30%) or OGTT (164, 70%), corresponding to a prevalence of 18%. The 70 indiviudals with cFBG ≥200 mg/dL, also all had HbA1c measurments >6.5%. Of the index tests evaluated, the urine glucose test strip had lower sensitivity (14.1%, 95% CI: 9.90-19.2) than cFBG (73.9%, 95% CI: 67.8-79.4), and HbA1c (75.2%, 95% CI: 69.2-80.6). All three tests offered high specificity (99.3%, 95% CI: 98.6-99.7; 96.8%, 95% CI: 95.5-97.8; and 98.5%, 95% CI: 97.5-99.1; respectively) (Table 2).The urine glucose test strip failed to identify 201 individuals with diabetes (false negatives) and falsely identified seven participants without diabetes (false positives). The 201 patients with diabetes who were not identified by the urine test had significantly lower venous FBG, lower 2 hr OGTT, and lower HbA1c compared to those correctly diagnosed, but were similar in other characteristics (Table 3). The seven false positive individuals had higher HbA1c, higher systolic BP, and higher proportion receiving treatment for hypertension than those with true negative results (Table 3).

The prevalence of diabetes (diagnosed by the composite reference standard) was significantly higher in participants who were 50 years of age or older compared to those under 50 years (24% vs. 9.6%, p<0.001); those with high BMI compared to those with normal BMI (22% vs. 17%, p=0.03); and those with greater waist circumference compared to those with normal waist (24% vs. 13%, p<0.001), but was the same in males and females (Table 4). The diagnostic accuracy of the urine glucose test strip was similar among subgroups of patients with various cofactors, with overlapping confidence intervals (Table 4). Additionally, multivariate and univariate logistic regression analyses also indicated that the diagnostic accuracy of the index test was not significantly impacted by these cofactors.

Discussion

Urine glucose test strips had much lower sensitivity than either cFBG or HbA1c, but all three tests offered high specificity. Patients who tested positive with the urine glucose test who were confirmed to have diabetes by the reference standard (true positives) had higher FBG, higher OGTT and higher HbA1c levels compared to the false negative group (urine test negative in patients with diabetes), suggesting that the urine glucose test may identify individuals with poor glycemic control. This suggests a subset of diabetes patients is being identified that may potentially be at higher risk of advancing complications or comorbidities, and who may benefit the most from further care [20]. In addition, testing for urine glucose was highly specific (99%), with positive LRs in the 20s, indicating that when positive, this test is highly indicative of diabetes.

The prevalence of diabetes in the MoPoTsyo population in Cambodia was 18%. This is much higher than the national prevalence for Cambodia, which is reported at 3.0%.[1] This may be due to the high proportion of individuals over 50 years of age in our study population, which could be explained by a participation bias towards those who were able to miss a day of work to attend a clinic visit. Additionally, our study took place in a rapidly changing urban population, which had a 2.4 times higher diabetes prevalence in the STEP survey, country report from 2010.[21]

A wide range of sensitivities for the urine glucose test strip has been reported, and its use remains controversial. A review in 2000 found six adequately designed studies that reported performance of urine test strips for glucose.[8] Among these, sensitivities in two reports of fasting patients were 16% and 35%; two using random samples found sensitivities of 18% and 64%; and three using postprandial and post-load measurements reported sensitivities between 39% and 48%. This review concluded that blood glucose measurements were preferred over urinary glucose or HbA1c, and particularly, postprandial over fasting measures. Another review found five studies reporting a range of sensitivity from 18% to 74% for urine glucose test strips.[7] The review concluded that urine glucose test strips are not sufficient for screening for diabetes.

This is one of the first studies to determine the prevalence of diabetes in Cambodia, and report on the screening accuracy of urine glucose test strips which are commonly used as screening tests in this setting. We used a prospective community-based design and had a large sample size with high participation rate. The study had several limitations. Firstly, we used a composite reference test and those with cFBG> 200 mg/dL were not evaluated by the OGTT. When evaluating the

index test of cFBG, the index test is included in the reference test, though at a different threshold, which can cause incorporation bias resulting in an inflated test accuracy. While OGTT is considered the gold standard reference test for assessing diagnostic accuracy, there has been some question of its performance. Two studies in China, each on more than 200 participants, found that the reproducibility of the OGTT was 56% [22] and 66% [23]. Though our choice of the reference standards, particularly OGTT, could have affected our study results, its use allows comparison of our results to those in a number of other studies. Second, the urine glucose test was self-administered and self reported. While this was pragmatic, and aligns with the practices at MoPoTyso and other clinical settings in Cambodia, errors in interpreting the test result could influence accuracy. We were not able to repeat this test when patients attended their clinic visit as they were fasting at the clinic visit, and thus their urine would not have been the random non-fasting urine test obtained at home. Third, we were not able to obtain hemoglobin levels (or test for hemoglobin variants) as these tests are not available in this setting, and hence cannot assess the impact of anemia or hemoglobinopathy on test performance.[24] Fourth, glucose test strip accuracy may be subject to effects of heat and humidity, we were not able to explore their possible impact on our results.

For clinicians working in settings similar to ours, the question is how useful is the urine glucose test as a screening or diagnostic test, and is it “better than nothing”? The low sensitivity certainly reduces the value of this test as a screening tool, but the high specificity means that positive tests can be used to rule in patients with diabetes, suggesting that urine glucose may have some diagnostic value in this setting. The false positive rate was extremely low, and only 7 patients without disease were identified as positive by urine glucose test strip. From a population

perspective, the value of a low cost, poorly sensitive yet highly specific test for diabetes is unclear in terms of balancing the opportunity to identify a subset of patients with less well controlled diabetes who would not have been identified otherwise, with the downside of a high false negative rate.[25]

Not surprisingly, usability parameters and cost make urine glucose test strips a highly desirable test in this and other low-resource settings.[9] Product attributes such as low complexity and infrastructure requirements, short time to results, and low participant burden greatly contribute to the acceptability and desirability of the screening tool. The large patient burden and the frequent inability to comply with fasting requirements reduce the feasibility of using OGTT or FBG tests. While HbA1c testing does not require fasting, current tests are too expensive for use in most low-income countries. The role of a poorly sensitive test like urine glucose in resource poor settings such as Cambodia is debatable, on the one hand the test will identify some patients previously undiagnosed, and assuming treatment can be initiated, reduce severity of complications from this disease. On the other hand, the test will miss the majority of patients with diabetes, thus risking a false reassurance, further postponement of diagnosis, and risking patient's respect for the health care system.

There may be strategies to improve the performance (particularly sensitivity) of the urine glucose test strip. First, using presence of risk factors such as high waist circumference or BMI, may increase the pretest probability of diabetes and lead to improved performance. In our study, the sensitivity of the UGTS among overweight men with high waist circumference was twice the overall sensitivity (29% vs. 14% respectiviely). Second, using random, postprandial, or glucose-

loaded measurements may be superior than fasting because the renal threshold for glucose is more often reached in non-fasting states.[8] Third, improving the limit of detection may be possible by modifications in the test strip itself, or improvement in the way it is read either manually (with trained users) or automatically (with electronic reading devices). Finally, increasing screening frequency may be feasible in low resource settings, if the urine glucose test strip truly does identify a smaller but more advanced fraction of diabetes patients.

Conclusion

Low cost, easy to use diabetes screening, diagnosis, and monitoring tools are essential for low-resource communities with minimal infrastructure. While the urine glucose test strip has some value as a screening test in these settings, its performance is far from optimal. Progress is urgently needed to improve the performance, availability, and access of essential testing technologies for diabetes.

List of abbreviations

urine glucose test strip (UGTS)

International Diabetes Federation (IDF)

non-communicable diseases (NCDs)

cardiovascular disease (CVD)

World Health Organisation (WHO)

capillary fasting blood glucose measurement (cFBG)

oral glucose tolerance test (OGTT)

positive predictive value (PPV)

negative predictive value (NPV)

positive likelihood ratio (LR+)

negative likelihood ratio (LR-)

confidence intervals (CI)

diabetes mellitus (DM)

Declarations

Ethical approval and consent to participate

The protocol was approved by the PATH Research Ethics Committee and the National Ethics Committee for Health Research (Cambodia Institutional Review Board). Informed consent was obtained from all participants.

Consent for publication

Not applicable.

Availability of data and material

The datasets used during the current study are available from the corresponding author on reasonable request.

Competing Interests

The authors declare that they have no competing interests.

Funding

This work was supported by a grant from Medtronic Foundation, and received additional support from PATH and the University of Washington Department of Family Medicine. The funding

source had no involvement in study design, data collection, data analysis, data interpretation, writing of the manuscript, or the decision to publish the results.

Authors contributions

MHP, SB, TN, HM and BW designed the study; MHP, SB, TN, and BW implemented the study; HLS, MT, HM, and BW analysed and interpreted the data; HLS, MHP, FD, MT, HM, and BW contributed to writing. All authors read and approved the final manuscript.

Acknowledgements

We would like to acknowledge the Borey Santhepheap community in Cambodia (Phnom Penh Municipality, District Dangkao, Commune Chom Chao) for participating in this study. We also acknowledge the input of Dr Annette Fitzpatrick and Dr Jim LoGerfo from the University of Washington.

Authors' information

Not applicable.

Tables

Table 1. Characteristics of included participants.

	Mean (SD) or % n=1289
Age, years	51.4 (14.9)
Female (%)	75.4
BMI[1]	23.2 (4.1)
High BMI (%)	30.5
Waist circumference above cutoff [2] (%)	46.1
Systolic blood pressure, mmHg	123.5 (20.6)
Diastolic blood pressure, mmHg	80.8 (12.1)
Elevated blood pressure (%)	12.9
Take treatment for high blood pressure (%)	8.2

[1] n=1288

[2] >90cm for men, >80cm for women. [19]

Table 2. Diagnostic accuracy of urine glucose test strip, capillary fasting glucose, and HbA1c determined by comparison with the composite reference standard (n=1289)[1].

	Urine glucose test strip positive	cFBG ≥126 mg/dL	HbA1c >6.5%
True positive (n)	33	173	176
False positive (n)	7	34	16
False negative (n)	201	61	58
True negative (n)	1048	1021	1039
True diabetes prevalence[2] (95%CI)	18%, 234/1289 (16, 20.4)		
Sensitivity (95% CI)	14.1 (9.90, 19.2)	73.9 (67.8, 79.4)	75.2 (69.2, 80.6)
Specificity (95% CI)	99.3 (98.6, 99.7)	96.8 (95.5, 97.8)	98.5 (97.5, 99.1)
Positive PV (95% CI)	82.5 (67.2, 92.7)	83.6 (77.8, 88.3)	91.7 (86.8, 95.2)
Negative PV (95% CI)	83.9 (81.7, 85.9)	94.4 (92.8, 95.7)	94.7 (93.2, 96.0)
Positive LR (95% CI)	21.3 (9.50, 47.5)	22.9 (16.3, 32.2)	49.6 (30.3, 81.1)
Negative LR (95% CI)	0.90 (0.80, 0.90)	0.30 (0.20, 0.30)	0.30 (0.20, 0.30)

[1] Excludes individuals taking diabetes treatment that day (n=6), did not fast before OGTT as instructed (n=5), or did not complete the OGTT (n=16).

[2] Composite reference standard: OGTT ≥200 mg/dL or cFBG ≥200 mg/dL. 70 patients with cFBG>=200 were not tested by OGTT.

Table 3. Diagnostic accuracy of the urine glucose test strip by patient characteristics.

	Diabetic[1]		Non-diabetic[1]	
Patient characteristic: Mean (SD) or %	**True Positive n=33**	**False Negative n=201**	**False Positive n=7**	**True Negative n=1048**
Age	57 (9.3)	58 (10.5)	56 (11.9)	50 (15.5)
Female (%)	81.8	74.6	85.7	75.3
Venous fasting blood glucose	**207 (75.3)**	**166 (73.2)**	95 (16.9)	90 (13.1)
Venous blood glucose 2 hrs after OGTT	**310 (60.8)**	**275 (62.2)**	115 (43.2)	120 (31.0)
Change in venous blood glucose during OGTT	160 (50.8)	146 (49.8)	20 (47.7)	30 (30.0)
HbA1c	**10 (2.3)**	**8 (2.4)**	**6 (0.7)**	**5 (0.5)**
BMI	24 (3.9)	24 (3.9)	26 (3.2)	23 (4.1)
High BMI (%)	33.3	36.8	57.1	29.0
Waist circumference above cutoff (%)	60.6	61.7	71.4	42.8
Systolic blood pressure	132 (24.9)	130 (20.6)	**146 (14.0)**	**122 (20.2)**
Diastolic blood pressure	85 (9.6)	84 (11.7)	87 (6.5)	80 (12.1)
Elevated blood pressure (%)	15.2	20.9	14.3	11.3
Take treatment for high blood pressure (%)	18.2	11.4	**28.6**	**7.1**
Total Cholesterol	242 (62.3)	227 (69.8)	240 (63.1)	213 (56.3)
Proteinuria (n=1116)[2] (%)	20.0	17.2	0	3.0
Albuminuria (%)	51.5	47.8	14.3	21.7
Abnormal albumin/creatinine ratio (%)	39.3	39.3	14.3	17.3

[1] Diagnosis by the composite reference standard: venous OGTT ≥200 mg/dL or cFBG ≥200 mg/dL. 70 patients with cFBG>=200 were not tested by OGTT.

[2] 4 missing values, 169 indeterminate measurements not included in analysis.

Bold = significantly different (p ≤ 0.05) by Student's t-test or chi-squared test.

Table 4. Diagnostic accuracy of urine glucose test strip by participant cofactors (n=1289) [1].

	Cofactors							
	Age		BMI [3]		Gender		Waist circumference [4]	
Results	**<50**	**≥50**	**<25**	**≥25**	**Male**	**Female**	**Normal**	**High**
Number of participants	531	758	895	393	317	972	691	598
True positive (n)	8	25	22	11	6	27	13	20
False positive (n)	3	4	3	4	1	6	2	5
False negative (n)	43	158	127	74	51	150	77	124
True negative (n)	477	571	743	304	259	789	599	449
True diabetes prevalence [2]	**9.6% (7.2, 12.4)**	**24% (21.0, 27.4)**	**17% (14.0, 19.3)**	**22% (18.0, 26.0)**	18% (14.0, 22.7)	18% (16.0, 20.8)	**13% (11.0, 15.8)**	**24% (21.0, 27.7)**
Sensitivity (95% CI)	15.7 (7.0, 28.6)	13.7 (9.0, 19.5)	14.8 (9.5, 21.5)	12.9 (6.6, 22.0)	10.5 (4.0, 21.5)	15.3 (10.3, 21.4)	14.4 (7.9, 23.4)	13.9 (8.7, 20.6)
Specificity (95% CI)	99.4 (98.2, 99.9)	99.3 (98.2, 99.8)	99.6 (98.8, 99.9)	98.7 (96.7, 99.6)	99.6 (97.9, 100)	99.2 (98.4, 99.7)	99.7 (98.8, 100)	98.9 (97.4, 99.6)
Positive PV (95% CI)	72.7 (39, 94.0)	86.2 (68.3, 96.1)	88 (68.8, 97.5)	73.3 (44.96, 92.2)	85.7 (42.1, 99.6)	81.8 (64.5, 93.0)	86.7 (59.5, 98.3)	80 (59.3, 93.2)
Negative PV (95% CI)	91.7 (89, 94)	78.3 (75.2, 81.3)	85.4 (82.9, 87.7)	80.4 76.1, 84.3)	83.5 (78.9, 87.5)	84 (81.5, 86.3)	88.6 (86, 90.9)	78.4 (74.8, 81.7)
Positive LR (95% CI)	25.1 (6.9, 91.6)	19.6 (6.9, 55.7)	36.7 (11.1, 121)	10.0 (3.3, 30.5)	27.4 (3.4, 223)	20.2 (8.5, 48.2)	43.4 (10.0, 189)	12.6 (4.8, 33)
Negative LR (95% CI)	0.8 (0.8, 1.0)	0.9 (0.8, 0.9)	0.9 (0.8, 0.9)	0.9 (0.8, 1.0)	0.9 (0.8, 1.0)	0.85 (0.80, 0.91)	0.86 (0.79, 0.94)	0.87 (0.82, 0.93)

[1] Excludes individuals taking diabetes treatment that day (n=6), did not fast before OGTT as instructed (n=5), or did not complete the OGTT (n=16).
[2] Composite reference standard: OGTT ≥200 mg/dL or cFBG ≥200 mg/dL. 70 patients with cFBG>=200 were not tested by OGTT.
[3] n=1288.
[4] High Waist circumference = >90cm for men, >80cm for women.[19]
Bold = significantly different (p ≤ 0.05), chi-squared test.

Figure legend

Figure 1: Study flow diagram.

Reference List

1. IDF Diabetes Atlas, 7th Edition. International Diabetes Federation . 2015. International Diabetes Federation.
Ref Type: Electronic Citation

2. Forbes JM, Cooper ME: **Mechanisms of diabetic complications.** *Physiol Rev* 2013, **93(1):** 137-188.

3. Tuomilehto J, Lindstrom J, Eriksson JG, Valle TT, Hamalainen H, Ilanne-Parikka P *et al.*: **Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance.** *N Engl J Med* 2001, **344:** 1343-1350.

4. Dyerberg J, Pedersen L, Aagaard O: **Evaluation of a dipstick test for glucose in urine.** *Clin Chem* 1976, **22:** 205-210.

5. van der Sande MA, Walraven GE, Bailey R, Rowley JT, Banya WA, Nyan OA *et al.*: **Is there a role for glycosuria testing in sub-Saharan Africa?** *Trop Med Int Health* 1999, **4:** 506-513.

6. Taal MW, Chertow GM, Marsden PA, Skorecki K, Yu ASL, Brenner BM: **Laboratory Assessment of Kidney Disease: Glomerular Filtration Rate, Urinalysis, and Proteinuria.** In *Brenner & Rector's The Kidney*. Philadelphia: Elsevier; 2012.

7. Wei OY, Teece S: **Best evidence topic report. Urine dipsticks in screening for diabetes mellitus.** *Emerg Med J* 2006, **23:** 138.

8. Engelgau MM, Narayan KM, Herman WH: **Screening for type 2 diabetes.** *Diabetes Care* 2000, **23:** 1563-1580.

9. World Health Organization. Screening for Type 2 Diabetes. Report of a World Health Organization and International Diabetes Federation meeting. 2003. Geneva, Switzerland, World Health Organization.
Ref Type: Report

10. van Pelt M: **Improving access to education and care in Cambodia.** *Diabetes Voice* 2009, **54**.

11. Taniguchi D, LoGerfo J, van PM, Mielcarek B, Huster K, Haider M *et al.*: **Evaluation of a multi-faceted diabetes care program including community-based peer educators in Takeo province, Cambodia, 2007-2013.** *PLoS One* 2017, **12:** e0181582.

12. van Olmen J, Eggermont N, van Pelt M, Hen H, de Man J, Schellevis F *et al.*: **Patient-centred innovation to ensure access to diabetes care in Cambodia: the case of MoPoTsyo.** *J Pharm Policy Pract* 2016, **9:** 1.

13. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L *et al.*: **STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies.** *BMJ* 2015, **351:** h5527.

14. World Health Organization. WHO STEPS Surveillance Manual. World Health Organization . 6-13-2008. World Health Organization. 4-30-2015.
Ref Type: Electronic Citation

15. Obuchowski NA: **Sample size calculations in studies of test accuracy.** *Stat Methods Med Res* 1998, **7:** 371-392.

16. King H, Keuky L, Seng S, Khun T, Roglic G, Pinget M: **Diabetes and associated disorders in Cambodia: two epidemiological surveys.** *Lancet* 2005, **366:** 1633-1639.

17. **Standards of Medical Care in Diabetes-2016 Abridged for Primary Care Providers.** *Clin Diabetes* 2016, **34:** 3-21.

18. Bartoli E, Fra GP, Carnevale Schianca GP: **The oral glucose tolerance test (OGTT) revisited.** *Eur J Intern Med* 2011, **22:** 8-12.

19. World Health Organization. Waist circumference and waist-hip ratio: report of a WHO expert consultation, Geneva, 8-11 December 2008. 2011. Geneva, Switzerland, World Health Organization.
Ref Type: Report

20. **The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The Diabetes Control and Complications Trial Research Group.** *N Engl J Med* 1993, **329:** 977-986.

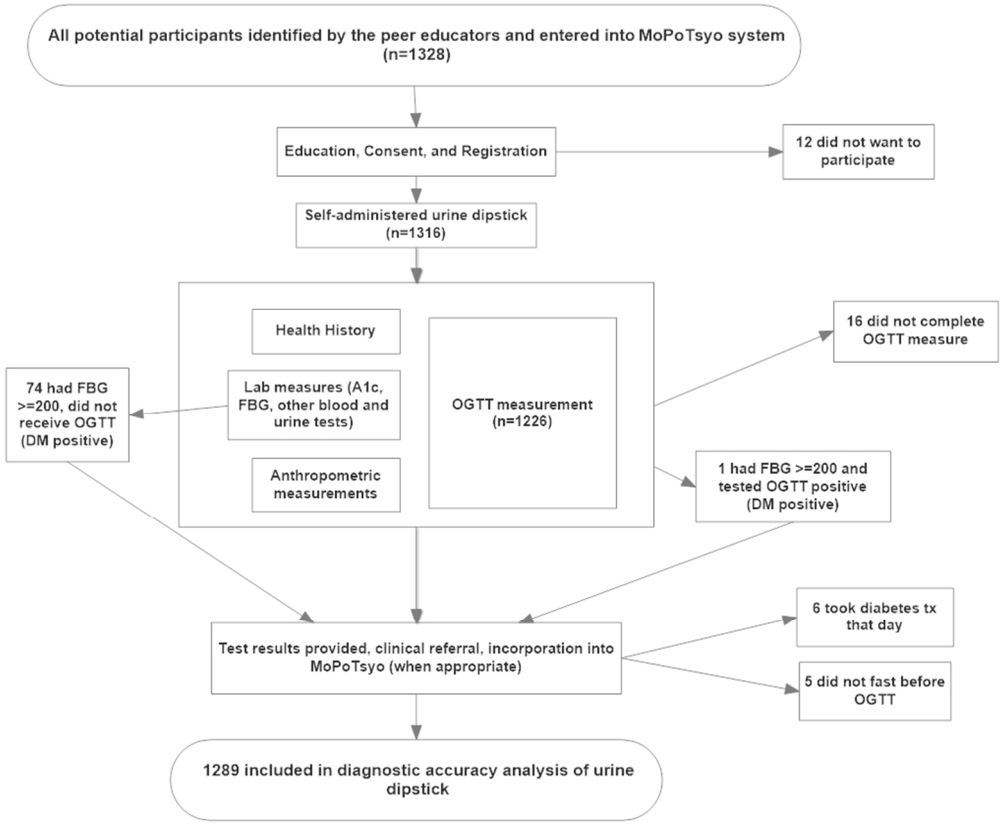
21. University of Health Sciences Cambodia, Ministry of Health Cambodia. Prevalence of non-communicable disease risk factors in Cambodia; STEPS Survey Country Report. 2010.
Ref Type: Report

22. Liu M, Pan CY, Jin MM, Su HY, Lu JM: **[The reproducibility and clinical significance of oral glucose tolerance test for abnormal glucose metabolism].** *Zhonghua Nei Ke Za Zhi* 2007, **46:** 1007-1010.

23. Ko GT, Chan JC, Woo J, Lau E, Yeung VT, Chow CC *et al.*: **The reproducibility and usefulness of the oral glucose tolerance test in screening for diabetes and other cardiovascular risk factors.** *Ann Clin Biochem* 1998, **35 (Pt 1):** 62-67.

24. English E, Idris I, Smith G, Dhatariya K, Kilpatrick ES, John WG: **The effect of anaemia and abnormalities of erythrocyte indices on HbA1c analysis: a systematic review.** *Diabetologia* 2015, **58:** 1409-1421.

25. Flessa S, Zembok A: **Costing of diabetes mellitus type II in Cambodia.** *Health Econ Rev* 2014, **4:** 24.



86x71mm (300 x 300 DPI)

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	1
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	3
	4	Study objectives and hypotheses	4
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	5
Participants	**6**	Eligibility criteria	5
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	5
	8	Where and when potentially eligible participants were identified (setting, location and dates)	5
	9	Whether participants formed a consecutive, random or convenience series	5
Test methods	**10a**	Index test, in sufficient detail to allow replication	5
	10b	Reference standard, in sufficient detail to allow replication	6
	11	Rationale for choosing the reference standard (if alternatives exist)	6
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	6
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	6
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	6
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	6
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	7
	15	How indeterminate index test or reference standard results were handled	7
	16	How missing data on the index test and reference standard were handled	7
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	7
	18	Intended sample size and how it was determined	6
RESULTS			
Participants	**19**	Flow of participants, using a diagram	17
	20	Baseline demographic and clinical characteristics of participants	15
	21a	Distribution of severity of disease in those with the target condition	15
	21b	Distribution of alternative diagnoses in those without the target condition	NA
	22	Time interval and any clinical interventions between index test and reference standard	5
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	15
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	15
	25	Any adverse events from performing the index test or the reference standard	NA
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	10
	27	Implications for practice, including the intended use and clinical role of the index test	11
OTHER INFORMATION			
	28	Registration number and name of registry	NA
	29	Where the full study protocol can be accessed	NA
	30	Sources of funding and other support; role of funders	14

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.

BMJ Open

Diagnostic accuracy of self-administered urine glucose test strips as a diabetes screening tool in a low-resource setting in Cambodia

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019924.R2
Article Type:	Research
Date Submitted by the Author:	16-Feb-2018
Complete List of Authors:	Storey, Helen; PATH, Diagnostics van Pelt, Maurits; MoPoTsyo Bun, Socheath; MoPoTsyo Daily, Frances; MoPoTsyo Neogi, Tina; PATH Thompson, Matthew; University of Washington, Department of Family Medicine McGuire, Helen; PATH, Noncommunicable Diseases Weigl, Bernhard; PATH
Primary Subject Heading:	Diabetes and endocrinology
Secondary Subject Heading:	Diagnostics, Global health
Keywords:	Low-resource settings, Diabetes, Diagnostics, Urine glucose test strip, Screening

Diagnostic accuracy of self-administered urine glucose test strips as a diabetes screening tool in a low-resource setting in Cambodia

Helen L. Storey; PATH, Seattle WA 98109, USA; hstorey@path.org

Maurits H. van Pelt; MoPoTsyo, Phnom Penh, Cambodia; mopotsyo@gmail.com

Socheath Bun; MoPoTsyo, Phnom Penh, Cambodia; mopo.medtrainer@gmail.com

Frances Daily; MoPoTsyo, Phnom Penh, Cambodia; fdaily@bigpond.com

Tina Neogi[1]; PATH, Seattle WA 98109, USA; tinaneogi@gmail.com

Matthew Thompson; Department of Family Medicine, University of Washington, Seattle WA 98195, USA; mjt@uw.edu

Helen McGuire; PATH, Washington DC 20001, USA; hmcguire@path.org

Bernhard H Weigl[2]; PATH, Seattle WA 98109, USA; bweigl@intven.com

* Corresponding author: Helen McGuire, PATH, PO Box 900922 Seattle WA 98109, hmcguire@path.org, 202-540-4360

[1] Present Address: UW Neighborhood Ravenna Clinic, Seattle WA 98105

[2] Present Address: Intellectual Ventures/Global Good, Bellevue WA 98005, USA

Abstract (word count: 287)

Objective: Screening for diabetes in low resource countries is a growing challenge, necessitating tests that are resource and context appropriate. The aim of this study was to determine the

diagnostic accuracy of a self-administered urine glucose test strip compared to alternative diabetes screening tools in a low resource setting of Cambodia.

Design: Prospective cross-sectional study

Setting: Members of the Borey Santhepheap community in Cambodia (Phnom Penh Municipality, District Dangkao, Commune Chom Chao).

Participants: All households on randomly selected streets were invited to participate, and adults at least 18 years of age living in the study area were eligible for inclusion.

Outcomes: The accuracy of self-administered urine glucose test strip positivity, HbA1c >6.5%, and capillary fasting blood glucose measurement ≥126 mg/dL were assessed against a composite reference standard of capillary fasting blood glucose measurement ≥200 mg/dL or venous blood glucose 2 hours after oral glucose tolerance test ≥200 mg/dL.

Results: Of the 1289 participants, 234 (18%) had diabetes based on either capillary fasting blood glucose measurement (74, 32%) or the oral glucose tolerance test (160, 68%). The urine glucose test strip was 14% sensitive and 99% specific, and failed to identify 201 individuals with diabetes, while falsely identifying 7 without diabetes. Those missed by the urine glucose test strip had lower venous fasting blood glucose, lower venous blood glucose 2 hours after oral glucose tolerance test, and lower HbA1c compared with those correctly diagnosed.

Conclusions: Low cost, easy to use diabetes tools are essential for low-resource communities with minimal infrastructure. While the urine glucose test strip may identify persons with diabetes that might otherwise go undiagnosed in these settings, its poor sensitivity cannot be ignored. The massive burden of diabetes in low-resource settings demands improvements in test technologies.

Keywords: Diabetes, Low-resource settings, Diagnostics, Urine glucose test strip, Screening,

Article Summary (word count: 2261)

Strengths and limitations of the study

- This is one of the first studies to determine the prevalence of diabetes and report on the screening accuracy of urine glucose test strips in Cambodia, which are commonly used as screening tests in this setting.
- We used a prospective community-based design and had a large sample size with high participation rate, though participation bias towards those able to miss a day of work to attend a clinic visit may still have been an issue.
- Use of a composite reference test and not evaluating those with capillary fasting blood glucose > 200 mg/dL by the oral glucose tolerance test, could have affected our study results, though the use of oral glucose tolerance test allows comparison of our results to those in a number of other studies.
- The urine glucose test was self-administered and self reported, which is pragmatic and aligns with the practices at MoPoTyso and other clinical settings in Cambodia, however errors in interpreting the test result could influence accuracy.

Background

According to the International Diabetes Federation, 415 million adults are living with diabetes globally, almost half of which are undiagnosed, and this number is expected to increase to 642 million by 2040.[1] As is the case for most non-communicable diseases, three quarters of those affected live in low- and middle-income countries. In Cambodia for example, there are an estimated 230,000 people with diabetes, who are at risk for the associated micro- and macrovascular complications of this disease, including cardiovascular disease.[1,2] Strategies to

reduce cardiovascular disease risk may also prevent and control diabetes, which would further reduce rates of eye, kidney, and neural damage due to diabetes complications.[3] To facilitate screening and monitoring for diabetes in these low- and middle-income countries, a low-cost, point-of-care diagnostic test that is resource and context appropriate is needed.

In low-resource settings, urine glucose test strips have been used as diabetes screening tools because they are inexpensive, noninvasive, and easy to use.[4,5] While these tests do not require fasting and are user friendly, they can only detect glucose after it has exceeded the threshold for reabsorption by the kidneys and appears in the urine. The reported threshold varies and is affected by kidney function.[6] Although their low sensitivity makes them inadequate for use as a screening tool,[7-9] the World Health Organisation acknowledges that they may have a place in low resource settings where other tests are not possible and the prevalence of undiagnosed diabetes may be high.[9] Currently many people are not diagnosed until severe complications develop. Although the sensitivity of the urine test delays diagnosis relative to other methods, it may provide an opportunity to reduce further advancement of complications.

MoPoTsyo, a nongovernmental organization, provides screening and care services to people with diabetes and hypertension in Cambodia through an innovative, community-based peer educator model.[10-12] MoPoTsyo uses urine glucose test strips issued in the community and self-administered by patients as the initial method of diabetes screening, which has allowed them to screen over 700,000 adults, followed by confirmation with blood glucose testing for those who have a positive urine test. The aim of this study was to determine the diagnostic accuracy of a self-administered urine glucose test strip compared to alternative diabetes screening tools in a

low resource setting of Cambodia. We also explored whether individuals with diabetes who were detected by urine glucose test strips differed in health status compared to those who were missed by this test but detected by blood glucose measurement. Greater understanding of the performance of this test by the MoPoTsyo program will help to inform its optimal use.

Methods

Study design and procedures

A prospective cross-sectional study was performed among members of the Borey Santhepheap community in Cambodia (Phnom Penh Municipality, District Dangkao, Commune Chom Chao) from November 2013 to October 2014. All households on randomly selected streets were invited to participate by a local peer educator, who described the study to all potential household members. Adults at least 18 years of age living in the study area were eligible for inclusion. Individuals were excluded if they had diabetes or hypertension or had taken medications for diabetes and/or high blood pressure in the last 30 days, had kidney disease, or had received dialysis. Written informed consent was obtained from all participants.The protocol was approved by the PATH Research Ethics Committee and the National Ethics Committee for Health Research (Cambodia Institutional Review Board). Study methods and results are reported in alignment with the 2015 STARD recommendations.[13]

After enrollment, all participants were screened for diabetes using a self-administered and self-reported urine glucose test strip (Sichuan Medicines and Health Products, Chengdu, China). Participants were taught how to use the test strip and read the results with assistance of a color chart, and were given several ways to report results to their peer educator. All participants were

then invited to attend the clinic following an 8-hour fast for laboratory confirmed tests for diabetes and associated co-morbid risk factors. Upon arriving at the clinic all participants provided a urine sample, a venous blood sample, and a finger stick blood sample for capillary fasting blood glucose measurement (cFBG) (On Call Plus glucometer, Acon Laboratories, San Diego, USA, https://www.aconlabs.com/us/glucose/on-call/plus-bgms/). If the cFBG was less than 200 mg/dL they were asked to consume a 75g oral glucose load for the oral glucose tolerance test (OGTT). The oral glucose load was ingested within 5 minutes of starting consumption, and two hours after ingestion, further venous blood and finger stick blood samples were obtained for glucose measurements. During the visit, a health history was completed based on the WHO STEPS surveillance questionnaire [14] and blood pressure measured by trained clinical staff using an ectronic device (Omron Corporation, Tokyo, Japan). All devices used in the study were owned and used previously by MoPoTsyo within the guidelines of the Cambodian Ministry of Health; none of the devices were investigational. Additional laboratory tests performed included HbA1c (DCA Vantage Analyzer, Siemens AG, Germany), serum creatinine, glucose, total cholesterol, high-density lipoprotein cholesterol, and triglycerides (Humalyzer 3000 Chemistry Analyzer, Human Diagnostics, Germany), spot urine creatinine, protein, and albumin tests (Combilyzer dipstick reader, Human Diagnostics, Germany).

A sample size of 1315 participants was calculated for a desired precision range of 10% and an estimated sensitivity and specificity of the urine glucose test strip of 21% and 90%, respectively, which is also sufficient for analysis of HbA1c, OGTT, and FBG as the test strip has the lowest performance. The sample size for the study was calculated based on Buderer's formula [15], accounting for a 3% drop-out rate and a 5% national prevalence of diabetes [16].

Data Analysis

The index tests of interest were a positive self-administered urine glucose test strip, HbA1c >6.5%, and cFBG ≥126 mg/dL. Diagnostic accuracy was assessed against a composite reference standard, which was cFBG ≥200 mg/dL, or venous blood glucose, 2 hours after OGTT ≥200 mg/dL.[17,18] If the participant's cFBG was >200 mg/dL, the patient was considered to have diabetes and an OGTT was not performed. Other measures were defined as follows: Overweight (BMI ≥25 or waist circumference >90cm for men or >80cm for women[19]), elevated blood pressure (systolic pressure ≥140mmHg or diastolic pressure ≥90mmHg), albuminuria (≥20 mg/L), and elevated albumin/creatinine ratio (≥30mg/g). We calculated sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+), negative LR (LR-),with 95% confidence intervals (CI).

Subgroup analyses were not prespecified, and therefore used to explore the performance of the urine glucose test strip in participants at increased risk for diabetes mellitus (DM), including age (>=50 years), BMI (>=25), gender, and waist circumference (>90cm for men or >80cm for women). Logistic regression analyses were also used to determine if the diagnostic accuracy of the index test was impacted by these clinical features. Prevalence of diabetes by subgroup was compared by chi-squared test. We also explored whether the individuals correctly classified by the urine glucose test strip had better or worse controlled diabetes than those misclassified by the test, as defined by various clinical and laboratory measures. Mean values of continuous variables were compared using Student's t-test while proportions of dichotomous values were compared using the chi-squared test. Data were analyzed using Stata/SE 13.1 (StataCorp LP, Texas, USA).

Results

Of 1328 eligible study subjects, 1316 participated in the study and 1289 were included in the analysis (Figure 1). Participants were excluded from the analysis if they did not complete the OGTT due to vomiting or other reasons (16), were not fasting prior to the clinic visit (5), or reported taking medication for diabetes that day (6). Of the analyzed participants, 75% (972/1289) were female, mean age was 51 years, 31% had high BMI, and 13% had elevated blood pressure, although only 8% were taking antihypertensive medications. Characteristics of the participants included in the analysis are presented in Table 1.

A total of 234 individuals had diabetes based on the composite reference standard of either cFBG (70, 30%) or OGTT (164, 70%), corresponding to a prevalence of 18%. The 70 indiviudals with cFBG ≥200 mg/dL, also all had HbA1c measurments >6.5%. Of the index tests evaluated, the urine glucose test strip had lower sensitivity (14.1%, 95% CI: 9.90-19.2) than cFBG (73.9%, 95% CI: 67.8-79.4), and HbA1c (75.2%, 95% CI: 69.2-80.6). All three tests offered high specificity (99.3%, 95% CI: 98.6-99.7; 96.8%, 95% CI: 95.5-97.8; and 98.5%, 95% CI: 97.5-99.1; respectively) (Table 2).The urine glucose test strip failed to identify 201 individuals with diabetes (false negatives) and falsely identified seven participants without diabetes (false positives). The 201 patients with diabetes who were not identified by the urine test had significantly lower venous FBG, lower 2 hr OGTT, and lower HbA1c compared to those correctly diagnosed, but were similar in other characteristics (Table 3). The seven false positive individuals had higher HbA1c, higher systolic BP, and higher proportion receiving treatment for hypertension than those with true negative results (Table 3).

The prevalence of diabetes (diagnosed by the composite reference standard) was significantly higher in participants who were 50 years of age or older compared to those under 50 years (24% vs. 9.6%); those with high BMI compared to those with normal BMI (22% vs. 17%); and those with greater waist circumference compared to those with normal waist (24% vs. 13%), but was the same in males and females (Table 4). The diagnostic accuracy of the urine glucose test strip was similar among subgroups of patients with various cofactors, with overlapping confidence intervals (Table 4). Additionally, multivariate and univariate logistic regression analyses also indicated that the diagnostic accuracy of the index test was not significantly impacted by these cofactors.

Discussion

Urine glucose test strips had much lower sensitivity than either cFBG or HbA1c, but all three tests offered high specificity. Patients who tested positive with the urine glucose test who were confirmed to have diabetes by the reference standard (true positives) had higher FBG, higher OGTT and higher HbA1c levels compared to the false negative group (urine test negative in patients with diabetes), suggesting that the urine glucose test may identify individuals with poor glycemic control. This suggests a subset of diabetes patients is being identified that may potentially be at higher risk of advancing complications or comorbidities, and who may benefit the most from further care [20]. In addition, testing for urine glucose was highly specific (99%), with positive LRs in the 20s, indicating that when positive, this test is highly indicative of diabetes.

The prevalence of diabetes in the MoPoTsyo population in Cambodia was 18%. This is much higher than the national prevalence for Cambodia, which is reported at 3.0%.[1] This may be due to the high proportion of individuals over 50 years of age in our study population, which could be explained by a participation bias towards those who were able to miss a day of work to attend a clinic visit. Additionally, our study took place in a rapidly changing urban population, which had a 2.4 times higher diabetes prevalence in the STEP survey, country report from 2010.[21]

A wide range of sensitivities for the urine glucose test strip has been reported, and its use remains controversial. A review in 2000 found six adequately designed studies that reported performance of urine test strips for glucose.[8] Among these, sensitivities in two reports of fasting patients were 16% and 35%; two using random samples found sensitivities of 18% and 64%; and three using postprandial and post-load measurements reported sensitivities between 39% and 48%. This review concluded that blood glucose measurements were preferred over urinary glucose or HbA1c, and particularly, postprandial over fasting measures. Another review found five studies reporting a range of sensitivity from 18% to 74% for urine glucose test strips.[7] The review concluded that urine glucose test strips are not sufficient for screening for diabetes.

This is one of the first studies to determine the prevalence of diabetes in Cambodia, and report on the screening accuracy of urine glucose test strips which are commonly used as screening tests in this setting. We used a prospective community-based design and had a large sample size with high participation rate. The study had several limitations. Firstly, we used a composite reference test and those with cFBG> 200 mg/dL were not evaluated by the OGTT. When evaluating the

index test of cFBG, the index test is included in the reference test, though at a different threshold. This can cause incorporation bias resulting in an inflated test accuracy. Here the three different index tests are included for comparison; however, the likely overestimation of diagnostic accuracy for cFBG is important to keep in mind. While OGTT is considered the gold standard reference test for assessing diagnostic accuracy, there has been some question of its performance. Two studies in China, each on more than 200 participants, found that the reproducibility of the OGTT was 56% [22] and 66% [23]. Though our choice of the reference standards, particularly OGTT, could have affected our study results, its use allows comparison of our results to those in a number of other studies. Second, the urine glucose test was self-administered and self reported. While this was pragmatic, and aligns with the practices at MoPoTyso and other clinical settings in Cambodia, errors in interpreting the test result could influence accuracy. We were not able to repeat this test when patients attended their clinic visit as they were fasting at the clinic visit, and thus their urine would not have been the random non-fasting urine test obtained at home. Third, we were not able to obtain hemoglobin levels (or test for hemoglobin variants) as these tests are not available in this setting, and hence cannot assess the impact of anemia or hemoglobinopathy on test performance.[24] Fourth, glucose test strip accuracy may be subject to effects of heat and humidity, we were not able to explore their possible impact on our results.

For clinicians working in settings similar to ours, the question is how useful is the urine glucose test as a screening or diagnostic test, and is it “better than nothing”? The low sensitivity certainly reduces the value of this test as a screening tool, but the high specificity means that positive tests can be used to rule in patients with diabetes, suggesting that urine glucose may have some diagnostic value in this setting. The false positive rate was extremely low, and only 7 patients

without disease were identified as positive by urine glucose test strip. From a population perspective, the value of a low cost, poorly sensitive yet highly specific test for diabetes is unclear in terms of balancing the opportunity to identify a subset of patients with less well controlled diabetes who would not have been identified otherwise, with the downside of a high false negative rate.[25]

Not surprisingly, usability parameters and cost make urine glucose test strips a highly desirable test in this and other low-resource settings.[9] Product attributes such as low complexity and infrastructure requirements, short time to results, and low participant burden greatly contribute to the acceptability and desirability of the screening tool. The large patient burden and the frequent inability to comply with fasting requirements reduce the feasibility of using OGTT or FBG tests. While HbA1c testing does not require fasting, current tests are too expensive for use in most low-income countries. The role of a poorly sensitive test like urine glucose in resource poor settings such as Cambodia is debatable, on the one hand the test will identify some patients previously undiagnosed, and assuming treatment can be initiated, reduce severity of complications from this disease. On the other hand, the test will miss the majority of patients with diabetes, thus risking a false reassurance, further postponement of diagnosis, and risking patient's respect for the health care system.

There may be strategies to improve the performance (particularly sensitivity) of the urine glucose test strip. First, using presence of risk factors such as high waist circumference or BMI, may increase the pretest probability of diabetes and lead to improved performance. In our study, the sensitivity of the urine glucose test strip among overweight men with high waist circumference

was twice the overall sensitivity (29% vs. 14% respectiviely). Second, using random, postprandial, or glucose-loaded measurements may be superior than fasting because the renal threshold for glucose is more often reached in non-fasting states.[8] Third, improving the limit of detection may be possible by modifications in the test strip itself, or improvement in the way it is read either manually (with trained users) or automatically (with electronic reading devices). Finally, increasing screening frequency may be feasible in low resource settings, if the urine glucose test strip truly does identify a smaller but more advanced fraction of diabetes patients.

Conclusion

Low cost, easy to use diabetes screening, diagnosis, and monitoring tools are essential for low-resource communities with minimal infrastructure. While the urine glucose test strip has some value as a screening test in these settings, its performance is far from optimal. Progress is urgently needed to improve the performance, availability, and access of essential testing technologies for diabetes.

List of abbreviations

urine glucose test strip (UGTS)

capillary fasting blood glucose measurement (cFBG)

oral glucose tolerance test (OGTT)

positive predictive value (PPV)

negative predictive value (NPV)

positive likelihood ratio (LR+)

negative likelihood ratio (LR-)

confidence intervals (CI)

diabetes mellitus (DM)

Declarations

Ethical approval and consent to participate

The protocol was approved by the PATH Research Ethics Committee and the National Ethics Committee for Health Research (Cambodia Institutional Review Board). Informed consent was obtained from all participants.

Consent for publication

Not applicable.

Availability of data and material

The datasets used during the current study are available from the corresponding author on reasonable request.

Competing Interests

The authors declare that they have no competing interests.

Funding

This work was supported by a grant from Medtronic Foundation, and received additional support from PATH and the University of Washington Department of Family Medicine. The funding source had no involvement in study design, data collection, data analysis, data interpretation, writing of the manuscript, or the decision to publish the results.

Authors contributions

MHP, SB, TN, HM and BW designed the study; MHP, SB, TN, and BW implemented the study; HLS, MT, HM, and BW analysed and interpreted the data; HLS, MHP, FD, MT, HM, and BW contributed to writing. All authors read and approved the final manuscript.

Acknowledgements

We would like to acknowledge the Borey Santhepheap community in Cambodia (Phnom Penh Municipality, District Dangkao, Commune Chom Chao) for participating in this study. We also acknowledge the input of Dr Annette Fitzpatrick and Dr Jim LoGerfo from the University of Washington.

Authors' information

Not applicable.

Tables

Table 1. Characteristics of included participants.

	Mean (SD) or % n=1289
Age, years	51.4 (14.9)
Female (%)	75.4
BMI[1]	23.2 (4.1)
High BMI (%)	30.5
Waist circumference above cutoff[2] (%)	46.1
Systolic blood pressure, mmHg	123.5 (20.6)
Diastolic blood pressure, mmHg	80.8 (12.1)
Elevated blood pressure (%)	12.9
Take treatment for high blood pressure (%)	8.2

[1] n=1288

[2] >90cm for men, >80cm for women. [19]

Table 2. Diagnostic accuracy of urine glucose test strip, capillary fasting glucose, and HbA1c determined by comparison with the composite reference standard (n=1289)[1].

	Urine glucose test strip positive	**cFBG ≥126 mg/dL**	**HbA1c >6.5%**
True positive (n)	33	173	176
False positive (n)	7	34	16
False negative (n)	201	61	58

True negative (n)	1048	1021	1039
True diabetes prevalence[2] (95%CI)	18%, 234/1289 (16, 20.4)		
Sensitivity (95% CI)	14.1 (9.90, 19.2)	73.9 (67.8, 79.4)	75.2 (69.2, 80.6)
Specificity (95% CI)	99.3 (98.6, 99.7)	96.8 (95.5, 97.8)	98.5 (97.5, 99.1)
Positive PV (95% CI)	82.5 (67.2, 92.7)	83.6 (77.8, 88.3)	91.7 (86.8, 95.2)
Negative PV (95% CI)	83.9 (81.7, 85.9)	94.4 (92.8, 95.7)	94.7 (93.2, 96.0)
Positive LR (95% CI)	21.3 (9.50, 47.5)	22.9 (16.3, 32.2)	49.6 (30.3, 81.1)
Negative LR (95% CI)	0.90 (0.80, 0.90)	0.30 (0.20, 0.30)	0.30 (0.20, 0.30)

[1] Excludes individuals taking diabetes treatment that day (n=6), did not fast before OGTT as instructed (n=5), or did not complete the OGTT (n=16).

[2] Composite reference standard: OGTT ≥200 mg/dL or cFBG ≥200 mg/dL. 70 patients with cFBG>=200 were not tested by OGTT.

Table 3. Diagnostic accuracy of the urine glucose test strip by patient characteristics.

	Diabetic[1]		**Non-diabetic[1]**	
Patient characteristic: Mean (SD) or %	**True Positive n=33**	**False Negative n=201**	**False Positive n=7**	**True Negative n=1048**
Age	57 (9.3)	58 (10.5)	56 (11.9)	50 (15.5)
Female (%)	81.8	74.6	85.7	75.3
Venous fasting blood glucose	**207 (75.3)**	**166 (73.2)**	95 (16.9)	90 (13.1)
Venous blood glucose 2 hrs after OGTT	**310 (60.8)**	**275 (62.2)**	115 (43.2)	120 (31.0)
Change in venous blood glucose during OGTT	160 (50.8)	146 (49.8)	20 (47.7)	30 (30.0)
HbA1c	**10 (2.3)**	**8 (2.4)**	**6 (0.7)**	**5 (0.5)**
BMI	24 (3.9)	24 (3.9)	26 (3.2)	23 (4.1)
High BMI (%)	33.3	36.8	57.1	29.0
Waist circumference above cutoff (%)	60.6	61.7	71.4	42.8
Systolic blood pressure	132 (24.9)	130 (20.6)	**146 (14.0)**	**122 (20.2)**
Diastolic blood pressure	85 (9.6)	84 (11.7)	87 (6.5)	80 (12.1)
Elevated blood pressure (%)	15.2	20.9	14.3	11.3
Take treatment for high blood pressure (%)	18.2	11.4	**28.6**	**7.1**
Total Cholesterol	242 (62.3)	227 (69.8)	240 (63.1)	213 (56.3)
Proteinuria (n=1116)[2] (%)	20.0	17.2	0	3.0
Albuminuria (%)	51.5	47.8	14.3	21.7
Abnormal albumin/creatinine ratio (%)	39.3	39.3	14.3	17.3

[1] Diagnosis by the composite reference standard: venous OGTT ≥200 mg/dL or cFBG ≥200 mg/dL. 70 patients with cFBG>=200 were not tested by OGTT.

[2] 4 missing values, 169 indeterminate measurements not included in analysis.

Bold = significantly different ($p \leq 0.05$) by Student's t-test or chi-squared test.

Table 4. Diagnostic accuracy of urine glucose test strip by participant cofactors (n=1289) [1].

Results	**Cofactors**

	Age		BMI [3]		Gender		Waist circumference [4]	
	<50	**≥50**	**<25**	**≥25**	**Male**	**Female**	**Normal**	**High**
Number of participants	531	758	895	393	317	972	691	598
True positive (n)	8	25	22	11	6	27	13	20
False positive (n)	3	4	3	4	1	6	2	5
False negative (n)	43	158	127	74	51	150	77	124
True negative (n)	477	571	743	304	259	789	599	449
True diabetes prevalence [2]	**9.6% (7.2, 12.4)**	**24% (21.0, 27.4)**	**17% (14.0, 19.3)**	**22% (18.0, 26.0)**	18% (14.0, 22.7)	18% (16.0, 20.8)	**13% (11.0, 15.8)**	**24% (21.0, 27.7)**
Sensitivity (95% CI)	15.7 (7.0, 28.6)	13.7 (9.0, 19.5)	14.8 (9.5, 21.5)	12.9 (6.6, 22.0)	10.5 (4.0, 21.5)	15.3 (10.3, 21.4)	14.4 (7.9, 23.4)	13.9 (8.7, 20.6)
Specificity (95% CI)	99.4 (98.2, 99.9)	99.3 (98.2, 99.8)	99.6 (98.8, 99.9)	98.7 (96.7, 99.6)	99.6 (97.9, 100)	99.2 (98.4, 99.7)	99.7 (98.8, 100)	98.9 (97.4, 99.6)
Positive PV (95% CI)	72.7 (39, 94.0)	86.2 (68.3, 96.1)	88 (68.8, 97.5)	73.3 (44.96, 92.2)	85.7 (42.1, 99.6)	81.8 (64.5, 93.0)	86.7 (59.5, 98.3)	80 (59.3, 93.2)
Negative PV (95% CI)	91.7 (89, 94)	78.3 (75.2, 81.3)	85.4 (82.9, 87.7)	80.4 76.1, 84.3)	83.5 (78.9, 87.5)	84 (81.5, 86.3)	88.6 (86, 90.9)	78.4 (74.8, 81.7)
Positive LR (95% CI)	25.1 (6.9, 91.6)	19.6 (6.9, 55.7)	36.7 (11.1, 121)	10.0 (3.3, 30.5)	27.4 (3.4, 223)	20.2 (8.5, 48.2)	43.4 (10.0, 189)	12.6 (4.8, 33)
Negative LR (95% CI)	0.8 (0.8, 1.0)	0.9 (0.8, 0.9)	0.9 (0.8, 0.9)	0.9 (0.8, 1.0)	0.9 (0.8, 1.0)	0.85 (0.80, 0.91)	0.86 (0.79, 0.94)	0.87 (0.82, 0.93)

[1] Excludes individuals taking diabetes treatment that day (n=6), did not fast before OGTT as instructed (n=5), or did not complete the OGTT (n=16).
[2] Composite reference standard: OGTT ≥200 mg/dL or cFBG ≥200 mg/dL. 70 patients with cFBG>=200 were not tested by OGTT.
[3] n=1288.
[4] High Waist circumference = >90cm for men, >80cm for women.[19]
Bold = significantly different (p ≤ 0.05), chi-squared test.

Figure legend

Figure 1: Study flow diagram.

Reference List

1. IDF Diabetes Atlas, 7th Edition. International Diabetes Federation . 2015. International Diabetes Federation.
Ref Type: Electronic Citation

2. Forbes JM, Cooper ME: **Mechanisms of diabetic complications.** *Physiol Rev* 2013, **93(1):** 137-188.

3. Tuomilehto J, Lindstrom J, Eriksson JG, Valle TT, Hamalainen H, Ilanne-Parikka P *et al.*: **Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance.** *N Engl J Med* 2001, **344:** 1343-1350.

4. Dyerberg J, Pedersen L, Aagaard O: **Evaluation of a dipstick test for glucose in urine.** *Clin Chem* 1976, **22:** 205-210.

5. van der Sande MA, Walraven GE, Bailey R, Rowley JT, Banya WA, Nyan OA *et al.*: **Is there a role for glycosuria testing in sub-Saharan Africa?** *Trop Med Int Health* 1999, **4:** 506-513.

6. Taal MW, Chertow GM, Marsden PA, Skorecki K, Yu ASL, Brenner BM: **Laboratory Assessment of Kidney Disease: Glomerular Filtration Rate, Urinalysis, and Proteinuria.** In *Brenner & Rector's The Kidney*. Philadelphia: Elsevier; 2012.

7. Wei OY, Teece S: **Best evidence topic report. Urine dipsticks in screening for diabetes mellitus.** *Emerg Med J* 2006, **23:** 138.

8. Engelgau MM, Narayan KM, Herman WH: **Screening for type 2 diabetes.** *Diabetes Care* 2000, **23:** 1563-1580.

9. World Health Organization. Screening for Type 2 Diabetes. Report of a World Health Organization and International Diabetes Federation meeting. 2003. Geneva, Switzerland, World Health Organization.
Ref Type: Report

10. van Pelt M: **Improving access to education and care in Cambodia.** *Diabetes Voice* 2009, **54**.

11. Taniguchi D, LoGerfo J, van PM, Mielcarek B, Huster K, Haider M *et al.*: **Evaluation of a multi-faceted diabetes care program including community-based peer educators in Takeo province, Cambodia, 2007-2013.** *PLoS One* 2017, **12:** e0181582.

12. van Olmen J, Eggermont N, van Pelt M, Hen H, de Man J, Schellevis F *et al.*: **Patient-centred innovation to ensure access to diabetes care in Cambodia: the case of MoPoTsyo.** *J Pharm Policy Pract* 2016, **9:** 1.

13. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L *et al.*: **STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies.** *BMJ* 2015, **351:** h5527.

14. World Health Organization. WHO STEPS Surveillance Manual. World Health Organization . 6-13-2008. World Health Organization. 4-30-2015.
Ref Type: Electronic Citation

15. Obuchowski NA: **Sample size calculations in studies of test accuracy.** *Stat Methods Med Res* 1998, **7:** 371-392.

16. King H, Keuky L, Seng S, Khun T, Roglic G, Pinget M: **Diabetes and associated disorders in Cambodia: two epidemiological surveys.** *Lancet* 2005, **366:** 1633-1639.

17. **Standards of Medical Care in Diabetes-2016 Abridged for Primary Care Providers.** *Clin Diabetes* 2016, **34:** 3-21.

18. Bartoli E, Fra GP, Carnevale Schianca GP: **The oral glucose tolerance test (OGTT) revisited.** *Eur J Intern Med* 2011, **22:** 8-12.

19. World Health Organization. Waist circumference and waist-hip ratio: report of a WHO expert consultation, Geneva, 8-11 December 2008. 2011. Geneva, Switzerland, World Health Organization.
Ref Type: Report

20. **The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The Diabetes Control and Complications Trial Research Group.** *N Engl J Med* 1993, **329:** 977-986.

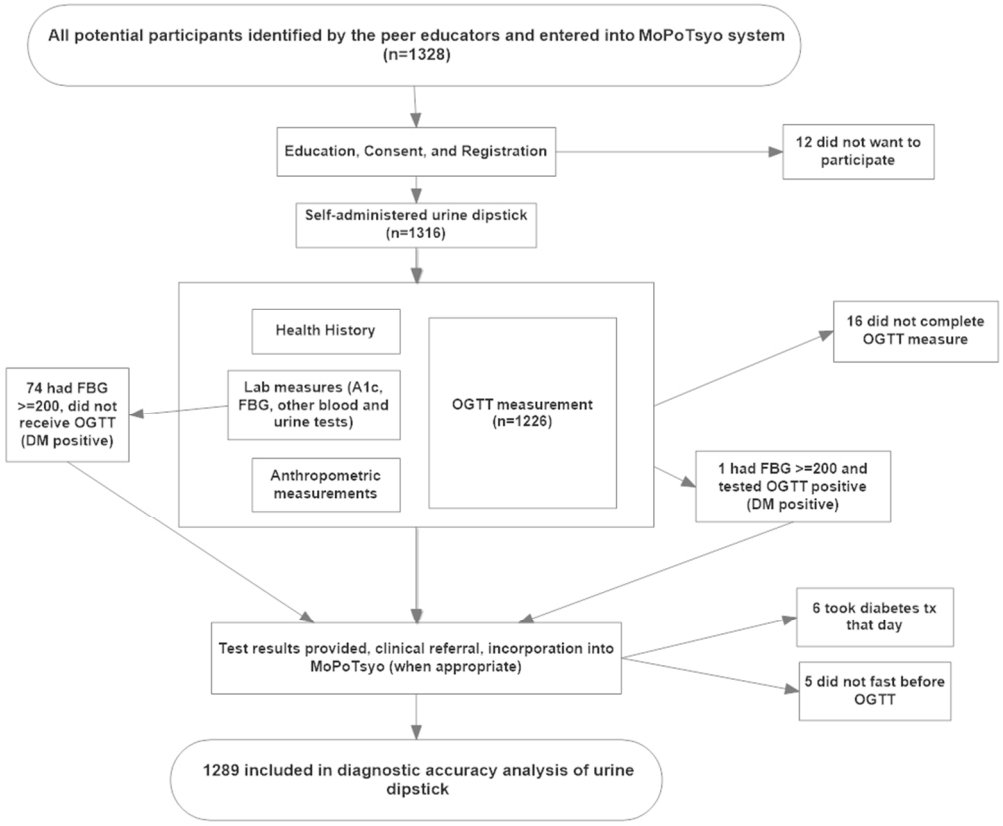
21. University of Health Sciences Cambodia, Ministry of Health Cambodia. Prevalence of non-communicable disease risk factors in Cambodia; STEPS Survey Country Report. 2010.
Ref Type: Report

22. Liu M, Pan CY, Jin MM, Su HY, Lu JM: **[The reproducibility and clinical significance of oral glucose tolerance test for abnormal glucose metabolism].** *Zhonghua Nei Ke Za Zhi* 2007, **46:** 1007-1010.

23. Ko GT, Chan JC, Woo J, Lau E, Yeung VT, Chow CC *et al*.: **The reproducibility and usefulness of the oral glucose tolerance test in screening for diabetes and other cardiovascular risk factors.** *Ann Clin Biochem* 1998, **35 (Pt 1):** 62-67.

24. English E, Idris I, Smith G, Dhatariya K, Kilpatrick ES, John WG: **The effect of anaemia and abnormalities of erythrocyte indices on HbA1c analysis: a systematic review.** *Diabetologia* 2015, **58:** 1409-1421.

25. Flessa S, Zembok A: **Costing of diabetes mellitus type II in Cambodia.** *Health Econ Rev* 2014, **4:** 24.



86x71mm (300 x 300 DPI)

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	1
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	3
	4	Study objectives and hypotheses	4
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	5
Participants	**6**	Eligibility criteria	5
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	5
	8	Where and when potentially eligible participants were identified (setting, location and dates)	5
	9	Whether participants formed a consecutive, random or convenience series	5
Test methods	**10a**	Index test, in sufficient detail to allow replication	5
	10b	Reference standard, in sufficient detail to allow replication	6
	11	Rationale for choosing the reference standard (if alternatives exist)	6
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	6
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	6
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	6
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	6
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	7
	15	How indeterminate index test or reference standard results were handled	7
	16	How missing data on the index test and reference standard were handled	7
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	7
	18	Intended sample size and how it was determined	6
RESULTS			
Participants	**19**	Flow of participants, using a diagram	17
	20	Baseline demographic and clinical characteristics of participants	15
	21a	Distribution of severity of disease in those with the target condition	15
	21b	Distribution of alternative diagnoses in those without the target condition	NA
	22	Time interval and any clinical interventions between index test and reference standard	5
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	15
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	15
	25	Any adverse events from performing the index test or the reference standard	NA
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	10
	27	Implications for practice, including the intended use and clinical role of the index test	11
OTHER INFORMATION			
	28	Registration number and name of registry	NA
	29	Where the full study protocol can be accessed	NA
	30	Sources of funding and other support; role of funders	14

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.