Journal of Microbiological Methods 177 (2020) 106047

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Clinical evaluation of the APAS® Independence: Automated imaging and interpretation of urine cultures using artificial intelligence with composite reference standard discrepant resolution

Lisa Brenton[a,*], Mary Jo Waters[a], Tyman Stanford[b,c], Steven Giglio[b,c]

[a] St. Vincent's Hospital, Melbourne, Australia
[b] LBT Innovations, Adelaide, Australia
[c] Clever Culture Systems, Switzerland

ARTICLE INFO

Keywords:
Automation
APAS
Artificial intelligence
Imaging
Discrepant

ABSTRACT

Background: This study reports the outcome of the first evaluation of the APAS® Independence for automated reading and preliminary interpretation of urine cultures in the routine clinical microbiology laboratory. In a 2-stage evaluation involving 3000 urine samples, two objectives were assessed; 1) the sensitivity and specificity of the APAS® Independence compared to microbiologists using colony enumeration as the primary determinant, and 2) the variability between microbiologists in enumerating bacterial cultures using traditional culture reading techniques, performed independently to APAS® Independence interpretation.

Methods: Routine urine samples received into the laboratory were processed and culture plates were interpreted by standard methodology and with the APAS® Independence. Results were compared using typical discrepant result resolution and with a composite reference standard, which provided an alternative assessment of performance.

Results: The significant growth sensitivity of the APAS® Independence was determined to be 0.919 with a 95% confidence interval of (0.879, 0.948), and the growth specificity was 0.877 with a 95% confidence interval of (0.827, 0.916). Variability between microbiologists was demonstrated with microbiologist bi-plate enumerations in agreement with the consensus 88.6% of the time.

Conclusion: The APAS® Independence appears to offer microbiology laboratories a mechanism to standardise the processing and assessment of urine cultures whilst augmenting the skills of specialist microbiology staff.

1. Introduction

The demand for pathology services continues to grow, driven largely by advances in medicine that allows for both greater life expectancy and the management of increasingly complex patients (Ledeboer and Dallas, 2014). At the same time, the ability of the microbiology laboratory to respond has become more challenging. Despite significant developments in automation, microbiology remains highly manual and labour-intensive and is reliant on the skills of microbiologists, especially in the area of interpretive assessment of microbial cultures. The introduction of artificial intelligence (AI) and machine learning technologies are promising to provide some solutions in this area (Croxatto et al., 2017; Faron et al., 2016a; Faron et al., 2016b).

The first AI-based system for clinical microbiology to receive FDA clearance was the APAS® Compact (Clever Culture Systems, Switzerland), an instrument used for the interpretation of routine urine samples cultured for 18 h on both Trypticase Soy Agar with 5% Sheep Blood and MacConkey Agar with crystal violet. A global clinical trial to validate the technology demonstrated a diagnostic sensitivity of 99.0% and a specificity of 84.5% (Glasson et al., 2017). The APAS® Compact, a Class II medical device, was a manual system that required the user to load single culture plates and allowed the release of an interpreted "No growth" result without user intervention. This was an important regulatory differentiator to other currently available digital plate reading systems that require microbiologist review of all images to release a result. However, using a single plate loading system in a routine clinical laboratory did not realise the impact of the technology.

The APAS® Independence (Clever Culture Systems, Switzerland) is a stand-alone in-vitro diagnostic instrument that fully automates culture plate imaging and interpretation and is an improvement on the APAS® Compact system by virtue of its high throughput and reduction in manual handling. The APAS® Independence differs from other imaging

* Corresponding author at: St Vincent's Melbourne, PO Box 2900, 41 Victoria Parade, Fitzroy VIC 3065, Australia.
E-mail address: Lisa.BRENTON@svha.org.au (L. Brenton).

https://doi.org/10.1016/j.mimet.2020.106047
Received 5 July 2020; Received in revised form 24 August 2020; Accepted 25 August 2020
Available online 11 September 2020

systems such as those found in the Kiestra (BD Life Sciences—Integrated Diagnostic Solutions, USA) and WASPLab (Copan Italia, Italy) in that it offers a plate reading function and does not include additional robotics to process or incubate specimens. Importantly, the core of the automated instrument (the way the images are captured) remains unchanged from the APAS® Compact and is substantially equivalent. The software used to interpret the images varies between the two systems and is tailored for specific use across different specimen types.

This paper describes the first laboratory evaluation of the APAS® Independence for automated reading and preliminary interpretation of routine urine cultures, using a bi-plate combination of Horse Blood Agar and UTI Brilliance Agar. This evaluation differs from previous observations of the APAS® Compact system as it was conducted within a routine clinical laboratory environment, using significantly different culture media and new interpretive software intended specifically for this media.

With a processing rate of 200 plates per hour, the APAS® Independence reads and interprets microbial cultures using proprietary classification algorithms for enumeration and classification, and in the case of urine samples, an expert decision system based on international reporting guidelines (England, 2017; Grabe et al., 2015; Kouri et al., 2000; McCarter et al., 2009). Whilst interpretative subtleties exist across these guidelines, their broad intent is that a bacterial load of 10^3 CFU/mL usually represents a low level of probability of urinary tract infection (UTI), and a colony count of $\geq 10^5$ CFU/mL indicates a high probability of UTI. A bacterial load of 10^4 CFU/mL requires more detailed interpretation in order to determine significance. It should be noted that there are additional indicators of significance aside from bacterial load when determining the presence of bacteriuria.

Using colony enumeration as the primary decision-making driver for this study, the instrument sorts urine culture plates into four designation categories, based on the likely significance of the culture; “Probable” ($\geq 10^5$ CFU/mL), “Review” (10^4 CFU/mL), “Doubtful” (10^3 CFU/mL), and “No Growth.” Cultures designated as “Probable” and “Review” (i.e. enumeration $\geq 10^4$ CFU/mL) require the expertise of microbiologists to further evaluate cultures and to perform identification and sensitivity testing as required. In contrast, “Doubtful” (typically skin and urogenital contamination, and a low likelihood of UTI) and “No Growth” designations require less skill, or no attention, and represent samples that could readily be reported through instrument function and interfacing.

Whilst total colony enumeration is generally the primary consideration in urine culture assessment, laboratory protocol-specific factors such as patient history and results of other laboratory tests, are frequently used as adjuncts to clinical interpretation and the generation of patient reports. The APAS® Independence has a bi-directional interface capability to support the input of these variables from a laboratory's information system (LIS), allowing the flagging of cultures for microbiologist review that, on enumeration and purity parameters alone, may not have been considered significant. When utilised, it provides an opportunity for users to customise reporting practices and the final clinical interpretation of the significance of cultures is not performed by the APAS® Independence alone.

In an evaluation involving 3000 urine samples, two objectives were assessed; 1) the reporting of the APAS® Independence compared to microbiologists performing traditional culture reading techniques, using colony enumeration as the primary determinant, and 2) the variability between microbiologists when enumerating urine cultures performed independently to APAS® Independence interpretation.

2. Methods

2.1. Standard urine processing methodology at St Vincent's Pathology

St Vincent's Pathology (SVP) provides services to St Vincent's Hospital, several affiliated private and specialist hospitals, general practitioners, and a large clinical trials business. Consequently, the microbiology laboratory receives a diverse range of specimens where urine samples requiring culture constitute approximately 50% of the total bacteriology workload. Up to 300 urine specimens are processed per day, and whilst a large proportion are considered routine, a range of complex samples are received from renal transplant recipients, catheterized inpatients and complicated oncology patients, for example.

Table 1
Definition and interpretive criteria of results by St Vincent's Pathology's standard workflow.

St Vincent's classification	Definition
No Growth[a]	No colony forming units detected
No significant growth (NSG)	$< 10^4$ CFU/mL
Significant growth (SIG)	$> 10^4$ CFU/mL

[a] No growth was included in the NSG group for analyses.

All urines received by SVP microbiology laboratory have automated chemistry and microscopy performed using the AUTION Hybrid AU-4050 System (ARKRAY, Inc., Japan). In addition, the majority of urines are cultured using a 1 μL calibrated loop onto a Horse Blood Agar (HBA)/Brilliance UTI chromogenic agar bi-plate (Thermo Fisher Scientific, Australia, product PP2249) (Fallon et al., 2002) and incubated aerobically at 35 °C for 18 h. Consistent with international reporting guidelines, interpretation of culture plates is initially determined by colony enumeration, and in the case of a bi-plate, the overall enumeration is calculated by taking the higher of the two agar colony counts, termed *amalgamated* enumeration. Table 1 provides a full description and definition of growth categories, where growth at $\geq 10^4$ CFU/mL is considered potentially significant (SIG), and growth at less than this level (or no growth) is generally considered not significant (NSG). All urine samples received for culture were included in this study. Samples did not contain preservatives.

2.2. APAS® Independence operation parameters

The APAS ® Independence was operated in accordance with manufacturer instructions, using Analysis Module AM007–1.0.0–27721. Following routine incubation, agar plates were loaded via the instrument's input module and the session was initiated to run. The instrument then mechanically transported the plates to an imaging station where plate images were acquired from both top- and bottom-lit light sources. Proprietary classification software was used to enumerate and analyse bacterial growth on each media of the bi-plate in order to determine the appropriate designation (Probable, Review, Doubtful, or No Growth) in the instrument's output section. This process took, on average, 18 s per bi-plate and occurred in real-time. As the primary objective of this study was to measure colony enumeration, plate designation categories offered by the system were not directly evaluated. The APAS® Independence was not interfaced to the LIS and results were not used for routine clinical reporting. For this study, growth $\geq 10^4$ CFU/mL was considered potentially significant, warranting microbiologist attention for further interpretation.

A schematic of workflow is detailed Fig. 1 and further detailed information on the system is available at www.cleverculturesystems.com.

2.3. APAS® Independence reporting compared to standard laboratory workflow

SVP performed this evaluation over a 5-week period. The aim was to assess the ability of the instrument to evaluate cultures compared to a microbiologist operating under current laboratory workflow and reporting conditions. In stage one a total of 881 cultures were analysed by the APAS® Independence and independently assessed and reported by microbiologists, according to the laboratory's routine protocols. To do

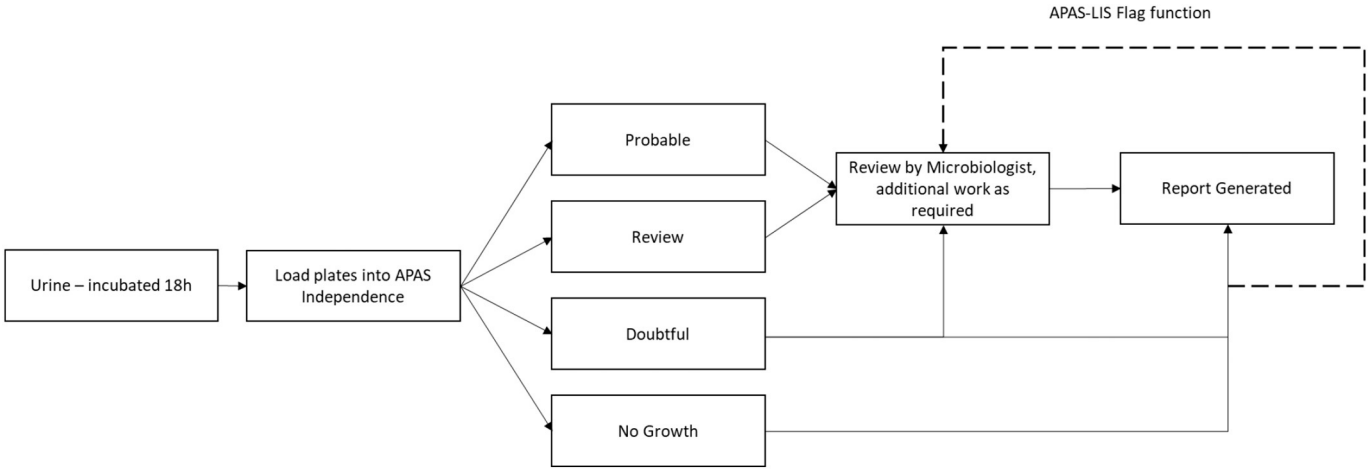

Fig. 1. Schematic representation of the APAS® Independence workflow within the laboratory. The APAS-LIS flag function provides the option for microbiologist review for No Growth or Doubtful plates under laboratory-defined conditions.

this, each side of the bi-plate was inoculated with 1 μL of well-mixed urine and manually streaked using a fishbone pattern. Plates were incubated for 18 h and were first analysed by the APAS® Independence and then by microbiologists. The two evaluations were performed using the same culture plates and the time between imaging and manual culture reading was typically less than one hour. The primary analysis variable in this study was bacterial growth enumeration of the bi-plate, classified as 0, 10^3, 10^4 and 10^5+ CFU/mL. The reporting microbiologists remained blinded to the APAS® Independence results.

The APAS® Independence calculated the enumeration for each agar of the bi-plate and assigned an amalgamated enumeration by taking the larger of these two values, in line with standard laboratory culture reading procedures performed by microbiologists. The final and routine culture results reported by SVP were extracted from the LIS. A direct comparison of results generated by SVP standard workflow and the APAS® Independence, is presented as a two by two confusion matrix in Table 2. Sensitivity and specificity were estimated with 95% confidence intervals using the Wilson score method (Newcombe, 1998). Sensitivity was defined as the probability of the APAS® Independence detecting significant growth (SIG, $\geq 10^4$ CFU/mL) where SVP determined significant growth, and specificity was defined as the probability of the APAS® Independence reporting non-significant growth where SVP determined non-significant growth (NSG, $< 10^4$ CFU/mL). Analysis was performed in R 3.5.1 (Team, 2017).

2.4. Investigation of microbiologist variability

Separate to the stage one comparison, a randomly selected subset of 480 plates (consisting of 107 No Growth plates, 151 plates with 10^3 CFU/mL, 148 with 10^4 CFU/mL, and 74 with $\geq 10^8$ CFU/mL) from the 2119 samples available for stage two, was used to assess the variability in culture interpretation between microbiologists. Three microbiologists assessed the enumeration of the APAS® Independence images independently of each other in a blinded manner. The microbiologists' enumeration values were then compared across all the microbiologists and for each of three pairwise combinations. A consensus enumeration value for each agar was calculated by taking the median of the ordinal enumeration categories of the microbiologists. Both percentage agreement and correlation using Kendall's τ_B (rank correlation) were calculated (Agresti, 2010).

Table 2
Sample classifications of St Vincent's Pathology's standard workflow (SVP) (rows) and APAS® Independence (columns) for the first stage of the evaluation. Entries in the table are the number of bi-plates with the given SVP and the APAS® Independence categorisations.

		APAS® Independence	
		NSG	SIG
SVP	NSG	342	90
	SIG	26	423

2.5. Composite reference standard (CRS) for discrepant result analysis

Discrepant resolution is commonplace in clinical microbiology, however inherently biased approaches investigating only samples that disagree are often used (Buss et al., 2015; Faron et al., 2016b; Faron et al., 2015; Harrington et al., 2015; Kost et al., 2017; Mashock et al., 2017). CRS methodology (Alonzo and Pepe, 1999a) avoids the bias inherent in typical discrepant resolution by not only sampling discrepant samples, but also sampling specimens in agreement, allowing and assessing the possibility that the samples in agreement may fall out of agreement with the use of an improved imperfect truth (such as the consensus of a panel of microbiologists). The CRS method requires a representative number of both discrepant and concordant samples, and the more samples analysed, the lower the width of the confidence intervals around revised CRS estimates. By not requiring all samples examined to have a composite reference standard calculated, CRS provides an efficient methodology whilst maintaining statistical rigour. The methodology to produce CRS-based sensitivity and specificity estimates with 95% confidence intervals is outlined in (Hawkins et al., 2001) using a probit transformation and the delta method (Cox, 1990). The data and code to perform the analysis in R are available at https://github.com/tystan/crs.

3. Results

3.1. APAS® Independence versus SVP

The data from the first stage of the evaluation demonstrated a high level of agreement (86.8%, 765/881; Table 2) between SVP and APAS® Independence enumerations. When there was disagreement, there was a larger probability the APAS® Independence would report the higher enumeration. Significant growth occurred on 449 plates as assessed by SVP; 423 of these plates were also assessed to have significant growth by APAS® Independence (Table 2). The significant growth sensitivity was estimated to be 0.942 with a 95% confidence interval of (0.917, 0.960). Similarly, 432 plates were assessed as having non-significant growth by SVP and 342 of these plates were also assessed to have non-

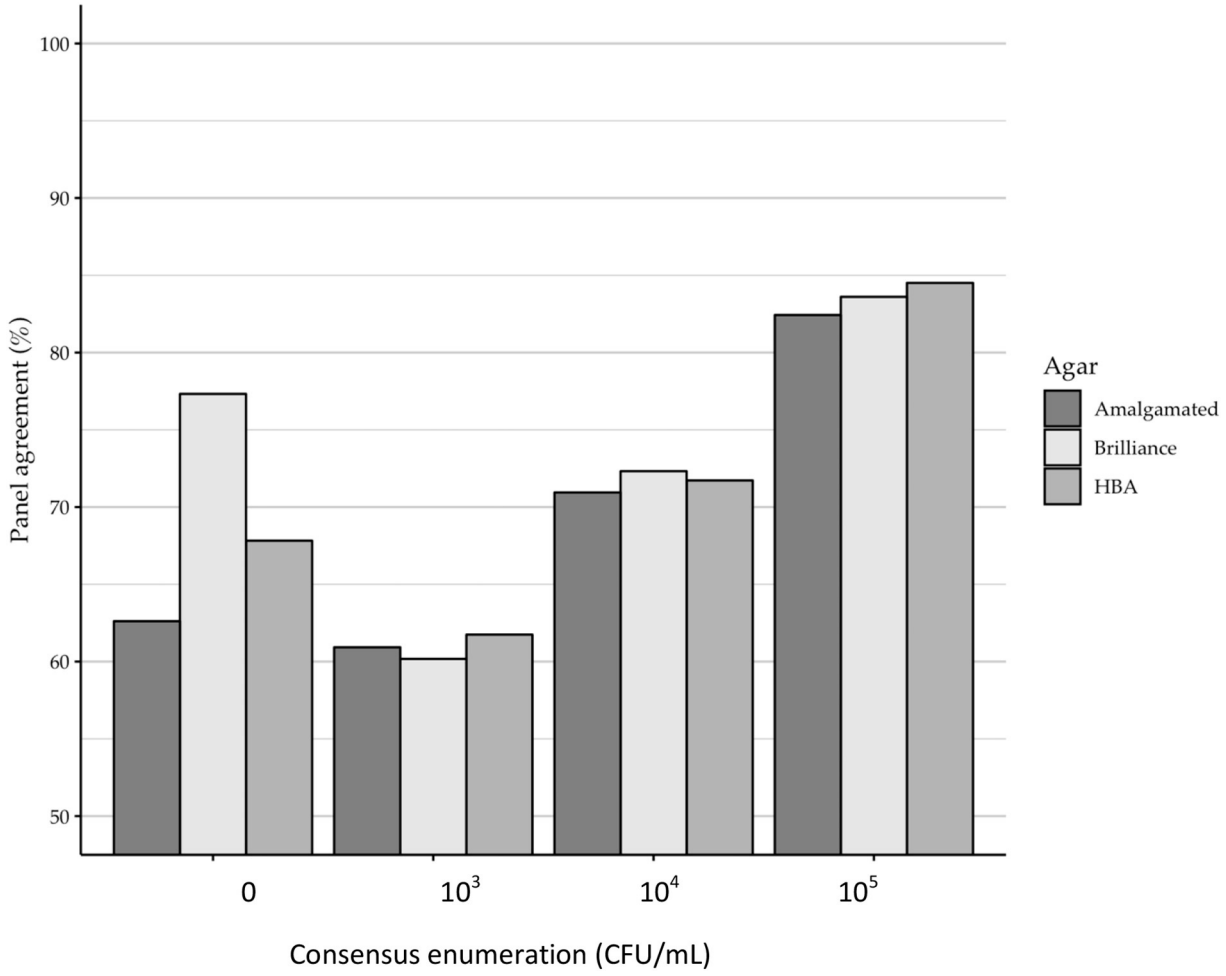

Fig. 2. Microbiologist panel agreement percentages by level of enumeration of each of two agars separately and for the amalgamated sample enumeration. Panel agreement is defined as all three microbiologists providing the same enumeration on the same plate image.

significant growth by APAS® Independence. The significant growth specificity was estimated to be 0.792 with a 95% confidence interval of (0.751, 0.827).

3.2. Microbiologist variability

Fig. 2 demonstrates the variability in urine culture enumeration across microbiologists. The lowest agreement on samples between all microbiologists occurred for consensus values of 10^3 CFU/mL for both agars (as low as 60.2% on Brilliance UTI agar (68/113)). As might be expected, the high growth (10^5+ CFU/mL) group showed the most concordance between all microbiologists for both agars. In almost a third of HBA cultures, one microbiologist reported some growth where the consensus was no growth (32.2%, 37/115). This occurred far less on Brilliance UTI agar (22.7%, 44/194).

Total agreement between all microbiologists for enumeration was achieved in 69.6% of samples on HBA (334/480) and 72.9% on Brilliance UTI agar (350/480). The amalgamated growth values were very similar to the HBA growth values because HBA usually exhibited more growth. However, microbiologist amalgamated enumerations in agreement with consensus was 88.6% (1276 /1440) which is slightly less than 94.5% agreement on HBA previously reported in (Glasson et al., 2016b). On HBA, 2 microbiologists demonstrated the most agreement with a percentage agreement of 82.7%. The same microbiologists had a similar level of agreement and correlation on Brilliance UTI agar. The percentage of agreement was slightly higher on Brilliance UTI agar for all pairwise comparisons of the microbiologist's enumeration values (data not shown), suggesting the reading of Brilliance UTI agar was perhaps simpler due to the presence of chromogenic reactions facilitating interpretation.

3.3. Composite reference standard

Fig. 3 demonstrates how the CRS method altered the initial achieved estimates for sensitivity and specificity. It is clear that the initial specificity result was an underestimate, as application of the CRS increased specificity considerably (by 0.085) to 0.877 (CRS specificity 95% CI: 0.827, 0.916). However, because of the nature of the CRS method, the sensitivity reduced from 0.942 to 0.919 (CRS sensitivity 95% CI: 0.879, 0.948). When considering the typical discrepant resolution method, labelled (c) *Discrepant resolution -typical*, both the sensitivity and specificity improved over the initial estimates, as would be expected.

4. Discussion

Using SVP as the truth status for plate assessment, the CRS significant growth sensitivity was determined to be 0.919 with a 95% confidence interval of (0.879, 0.948), whilst the CRS significant growth specificity was determined to be 0.877 with a 95% confidence interval of (0.827, 0.916). Where there was disagreement (n = 26, SVP SIG and APAS® Independence NSG), image review determined that twenty-four samples had urogenital/ skin contaminants at levels marginally above the 10^4 CFU/mL threshold. These were reported as mixed urogenital and skin contamination by the laboratory, but as < 10^4 CFU/mL by the

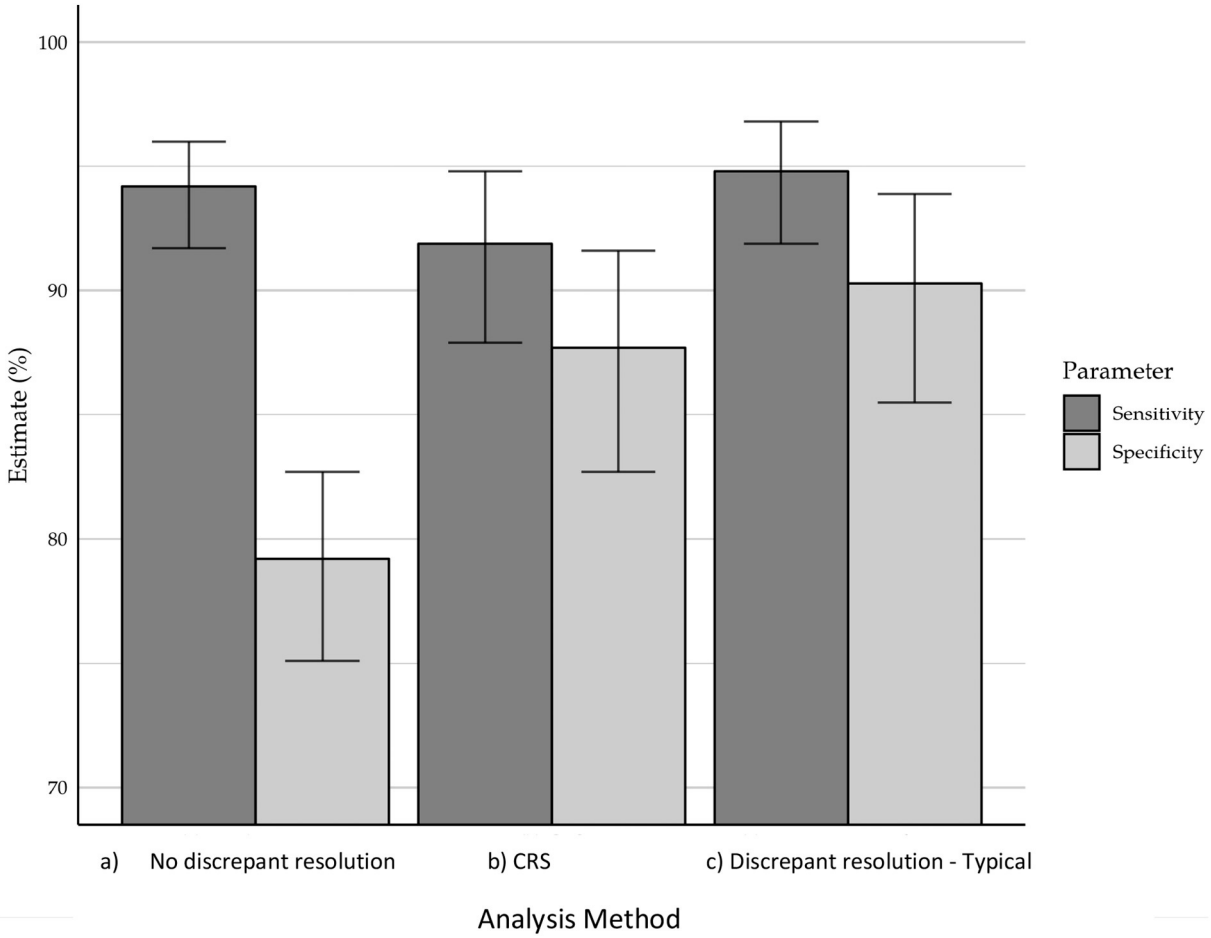

Fig. 3. Sensitivity and specificity estimates with 95% confidence intervals as vertical bands for the three analysis methods of estimation.

APAS® Independence. According to the study criteria, and based solely on bacterial enumeration, these results are considered false negatives for the APAS® Independence. However, this was not reflected in the final laboratory result of urogenital contamination, i.e. non-significant growth. The two remaining discrepant samples contained slow-growing alpha-haemolytic streptococci, presenting as hazy growth after 18 h incubation. Of these, one sample would have flagged for review/re-incubation by virtue of a raised white blood cell count and utilisation of a bi-directional interface to re-route the plate from a NSG classification to a SIG classification. A "flag" is feature of the APAS® Independence that can be used in conjunction with other sample parameters and patient demographics from the LIS to redirect culture classifications on the instrument in real-time. In both cases, the streptococcal isolates were reported by the laboratory as being of uncertain significance due to absence of any clinically relevant information. However this will not always be the case, and the limitation described here will require complementary laboratory procedures to minimise impact (Glasson et al., 2016a). Whilst the sensitivity estimate of 0.919 on enumeration alone may appear to be less than ideal, the limitations described in this study do not practically translate to a fully integrated instrument with LIS connectivity and the use of flags as described, which would predictably increase true clinically sensitivity. Assuming full LIS integration and the use of custom flags to promote those samples that are potentially clinically significant based on additional information, only one clinically significant case from 449 samples would have been misreported by the APAS® Independence, with a resulting sensitivity estimate of 0.998 (448/449 SIG samples, Table 2).

Lack of full integration of the APAS® Independence with the LIS was therefore a limitation of this study and such integration would additionally have allowed quantitative assessment of efficiencies attributable to automated plate reading and releasing of reports. This highlights the necessity to interrogate new technologies thoroughly in terms of both their analytical performance, and their true clinical performance with full laboratory integration.

Direct comparison to performance of previous studies using the APAS® Compact with Trypticase Soy Agar with 5% Sheep Blood and MacConkey Agar with crystal violet agars is not possible because the media used in this evaluation are different, but we note that the specificity in this study is greater while the raw sensitivity estimate is reduced. This may be a function of the media which poses some detection challenges to the interpretive software in the context of pin-point colonies on the agar. This is also likely to be a function of the way the results were analysed, as the previous studies measured sensitivity and specificity based on the presence or absence of growth, while this study uses bacterial enumeration as the primary driver for determining the likely significance of a urine culture. Whilst it is generally accepted that the bacterial load of urinary tract infections is $> 10^5$ CFU/mL, there are often exceptions to this, both clinical and laboratory-specific, that add a further study limitation.

Although human assessment of bacterial cultures remains the gold standard, this evaluation has shown its subjective and variable nature, which is surprisingly under-reported in the literature (Glasson et al., 2016b). The APAS® Independence allows for the standardisation of culture reading through the use of standardised interpretive software. In this study, microbiologist variability was assessed using electronic images of cultures and it may be that the low level of agreement at values of 10^3 CFU/mL was due, in part, to the misclassification of agar flecks, dust, and/or probable artefacts on images as colonies (2). With

any digital plate reading systems, ensuring appropriate training and competency assessment is especially important, especially in the absence of the ability to manipulate angles of light on the plate to determine form and shape. In the case of the APAS® Independence, efforts are made to computationally remove these common artefacts, and their classification as background is a key difference between the APAS® Independence and other digital plate reading systems that currently exploit differences in images at incubation points.

Despite the established and authoritative recommendations against typical discrepant resolution (Alonzo and Pepe, 1999b, Hadgu Hadgu, 1996, Hadgu, 2000, Hawkins et al., 2001, McAdam, 2000, McAdam, 2017, Miller, 1998), methods interrogating only those results that do not agree remains in widespread use (Buss et al., 2015; Faron et al., 2016b; Faron et al., 2015; Harrington et al., 2015; Kost et al., 2017; Mashock et al., 2017). An alternative, and apparently superior method, CRS, overcomes the shortcomings of an imperfect truth when evaluating a new test or screening method. CRS takes the classification of plates by a new test and an imperfect truth, and then improves the estimates of the test's performance by subsampling and re-assessing the plates in a manner that performs as an improved imperfect truth. CRS applies this methodology to both concordant and discrepant samples, not just discrepant samples, and therefore removes bias.

CRS provides a defensible method to reduce the effect of individual microbiologist variability unfairly altering performance estimates. These data demonstrate that more concordance was achieved when examining a subsample of plates with a panel of microbiologists, reducing variability in the truth status of plates. However, this does not necessarily equate to improved sensitivity and specificity estimates as seen here (specificity increased from 0.792 by 0.085, but sensitivity reduced from 0.942 by 0.023), while typical discrepant analysis did improve sensitivity and specificity as expected.

The use of AI in diagnostics is rapidly advancing, and when implementing systems to augment human skill and knowledge, there must be a concerted effort to truly understand the technology and its limitations, on its own and when combined with adjunct technologies or systems. When evaluating these new technologies, simply continuing typical analyses and processes is not applicable and not statistically defensible. It is a timely reminder to examine, with scientific and statistical rigour, representative samples from across the entire sample set to ensure that the true performance is understood and to confirm that the technology will be fit for diagnostic purpose. Simply setting typical performance measurements (such as sensitivity and specificity) may not be enough. It is also important to consider usability across staff and to ensure engagement with these stakeholders is given the appropriate focus and weight in the decision-making process for adopting new technologies.

In routine laboratory testing where approximately 60–70% of urine cultures return results of no growth or no significant growth, there exists a potential for significant reduction in microbiologist time to read and report these cultures using the APAS® Independence. This is primarily due to the triaging function of the instrument, and the fact that the instrument provides a report to the laboratory information system that can facilitate the automatic reporting of samples that yield results of no growth and no significant growth (NSG), requiring no microbiologist input and a decreased time to reporting. This opens the potential for workflow redesign and redirection of skilled resources to other areas of the laboratory.

The APAS® Independence adds to the suite of options from other manufacturers in microbiology. This includes solutions for specimen processing, for example, WASP (Copan Italia, Italy), InoqulA (BD Life Sciences - Integrated Diagnostic Solutions, USA), and AutoPlak (Beckman Coulter, USA) as well as automated incubation and imaging of samples (ReadA - BD Life Sciences - Integrated Diagnostic Solutions, USA) without any assisted artificial intelligence. Combined, the concept of workstation automation allows users to customise and prioritise system integrations for individual laboratories, and as fiscal conditions allow.

When evaluating new technologies with an evidence- and risk-based approach, clinical sensitivity is a key determinant for procurement and implementation. In this study, the clinical sensitivity availed by full implementation options was 0.998 which indicates the risk of false negatives is very low. However, other strategic considerations are also fundamental and were important considerations for the authors. As such, the APAS® Independence offers microbiology laboratories a mechanism to augment the skills of specialist microbiology staff, reduce workload and manual handling, and increase efficiencies and standardisation in the processing and assessment of urine cultures, as well as addressing a growing gap in the availability of skilled workers.

Declaration of Competing Interest

Tyman Stanford and Steven Giglio were employees of LBT Innovations and Clever Culture Systems at the time of writing this manuscript.

References

Agresti, A., 2010. Analysis of Ordinal Categorical Data. John Wiley & Sons.

Alonzo, T.A., Pepe, M.S., 1999a. Using a combination of reference tests to assess the accuracy of a new diagnostic test. Stat. Med. 18, 2987–3003.

Alonzo, T.A., Pepe, M.S., 1999b. Using a combination of reference tests to assess the accuracy of a new diagnostic test. Stat. Med. 18, 2987–3003.

Buss, S.N., Leber, A., Chapin, K., Fey, P.D., Bankowski, M.J., Jones, M.K., Rogatcheva, M., Kanack, K.J., Bourzac, K.M., 2015. Multicenter evaluation of the BioFire FilmArray gastrointestinal panel for etiologic diagnosis of infectious gastroenteritis. J. Clin. Microbiol. 53, 915–925.

Cox, C., 1990. Fieller's theorem, the likelihood and the delta method. Biometrics 709–718.

Croxatto, A., Marcelpoil, R., Orny, C., Morel, D., Prod'hom, G., Greub, G., 2017. Towards automated detection, semi-quantification and identification of microbial growth in clinical bacteriology: a proof of concept. Biom. J. 40, 317–328.

England, P.H., 2017. Investigation of Urine UK STANDARDS for Microbiology Investigations. B41 Issue 8.2.

Fallon, D., Andrews, N., Frodsham, D., Gee, B., Howe, S., Iliffe, A., Nye, K., Warren, R., 2002. A comparison of the performance of cystine lactose electrolyte deficient (CLED) agar with Oxoid chromogenic urinary tract infection (CUTI) medium for the isolation and presumptive identification of organisms from urine. J. Clin. Pathol. 55, 524–529.

Faron, M.L., Ledeboer, N.A., Granato, P., Daly, J.A., Pierce, K., Pancholi, P., Uphoff, T.S., Buchan, B.W., 2015. Detection of group a streptococcus in pharyngeal swab specimens by use of the AmpliVue GAS isothermal helicase-dependent amplification assay. J. Clin. Microbiol. 53, 2365–2367.

Faron, M.L., Buchan, B.W., Coon, C., Liebregts, T., Van Bree, A., Jansz, A.R., Soucy, G., Korver, J., Ledeboer, N.A., 2016a. Automatic digital analysis of chromogenic media for vancomycin-resistant-enterococcus screens using copan WASPLab. J. Clin. Microbiol. 54, 2464–2469.

Faron, M.L., Buchan, B.W., Vismara, C., Lacchini, C., Bielli, A., Gesu, G., Liebregts, T., Van Bree, A., Jansz, A., Soucy, G., Korver, J., Ledeboer, N.A., 2016b. Automated scoring of chromogenic media for detection of methicillin-resistant Staphylococcus aureus by use of WASPLab image analysis software. J. Clin. Microbiol. 54, 620–624.

Glasson, J., Hill, R., Summerford, M., Giglio, S., 2016a. Evaluation of an image analysis device (APAS) for screening urine cultures. J. Clin. Microbiol. 54, 300–304.

Glasson, J., Hill, R., Summerford, M., Giglio, S., 2016b. Observations on variations in manual Reading of cultures. J. Clin. Microbiol. 54, 2841.

Glasson, J., Hill, R., Summerford, M., Olden, D., Papadopoulos, F., Young, S., Giglio, S., 2017. Multicenter evaluation of an image analysis device (APAS): comparison between digital image and traditional plate reading using urine cultures. Ann. Lab. Med. 37, 499–504.

Grabe, M., Bartoletti, R., Bjerklund-Johansen, T., 2015. Guidelines on urological infections. EAU 2015.

Hadgu, A., 1996. The discrepancy in discrepant analysis. Lancet 348, 592–593.

Hadgu, A., 2000. Discrepant analysis is an inappropriate and unscientific method. J. Clin. Microbiol. 38, 4301–4302.

Harrington, S.M., Buchan, B.W., Doern, C., Fader, R., Ferraro, M.J., Pillai, D.R., Rychert, J., Doyle, L., Lainesse, A., Karchmer, T., Mortensen, J.E., 2015. Multicenter evaluation of the BD max enteric bacterial panel PCR assay for rapid detection of Salmonella spp., Shigella spp., campylobacter spp. (C. Jejuni and C. coli), and Shiga toxin 1 and 2 genes. J. Clin. Microbiol. 53, 1639–1647.

Hawkins, D.M., Garrett, J.A., Stephenson, B., 2001. Some issues in resolution of diagnostic tests using an imperfect gold standard. Stat. Med. 20, 1987–2001.

Kost, K., Yi, J., Rogers, B., Jerris, R., 2017. Comparison of clinical methods for detecting carbapenem-resistant Enterobacteriaceae. Pract. Lab. Med. 8, 18–25.

Kouri, T.T., Gant, V.A., Fogazzi, G.B., Hofmann, W., Hallander, H.O., Guder, W.G., 2000. Towards European urinalysis guidelines: introduction of a project under European Confederation of Laboratory Medicine. Clin. Chim. Acta 297, 305–311.

Ledeboer, N.A., Dallas, S.D., 2014. The automated clinical microbiology laboratory: fact

or fantasy? J. Clin. Microbiol. 52, 3140–3146.
Mashock, M.J., Faron, M.L., Buchan, B.W., Ledeboer, N.A., 2017. Evaluation of Copan FecalSwab as specimen type for use in Xpert C. difficile assay. J. Clin. Microbiol. 55, 3123–3129.
Mcadam, A.J., 2000. Discrepant analysis: how can we test a test? J. Clin. Microbiol. 38, 2027–2029.
Mcadam, A.J., 2017. Discrepant analysis and bias: a micro-comic strip. J. Clin. Microbiol. 55, 2878–2879.
Mccarter, Y., Burd, E., Hall, G., Zervos, M., 2009. Cumitech 2C: Laboratory Diagnosis of Urinary Tract Infections. American Society for Microbiology, Washington, DC.
Miller, W.C., 1998. Bias in discrepant analysis: when two wrongs don't make a right. J. Clin. Epidemiol. 51, 219–231.
Newcombe, R.G., 1998. Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat. Med. 17, 857–872.
Team, R.C., 2017. R: A Language and Environment for Statistical Computing, 3.4.3 ed. R Foundation for Statistical Computing, Vienna, Austria.