In February 2017, the Australian College for Infection Prevention and Control and the Australasian Society for Ultrasound in Medicine (ASUM) released joint guidelines for the reprocessing of ultrasound transducers. The document was developed in response to the results of an Australasian ultrasound-specific survey to determine current understanding of infection prevention and control in ultrasound practice. The additional goal was to increase awareness, via multiple published studies, for the potential of cross-contamination and infection transmission in ultrasound practice.

The ultrasound unit as a whole may be a vector for the transmission of potential pathogens to patients and staff. High-level disinfection for the reprocessing of intracavity transducers is accepted by most practitioners as a necessary adjunct to all invasive ultrasound examinations; however, the Australasian survey results indicate a lack of understanding for the need for correct low-level disinfection (LLD) of all scanning-related equipment after every use.

Bacterial contamination can be present on not only the transducer but also the keyboard, transducer connectors, gel bottles, and machine handles. Contamination cannot be excluded by visual inspection alone, with one study showing that only 51% of blood-contaminated samples were visibly stained and a second study showing that 23% of external transducers had bacterial contamination. Early disinfection could be reduced, leading to the possibility of persistence of active virus or bacteria.

The literature reveals multiple studies reporting contamination within the ultrasound unit involving external transducers. Early
reports were generally written around the risks of nosocomial infection arising from poorly or uncleaned ultrasound devices. More recent articles have investigated incidents of bacterial contamination, evidence of cross infection, methods for cleaning and decontamination, and even device degradation through inappropriate cleaning and decontamination. Each of these factors concomitantly lifts the risks associated with iatrogenic infection where the ultrasound probe could act as the critical fomite. A variety of bacterial pathogens and even environmental spore formers have been identified and associated with human infection, including unusual opportunistic pathogens such as Acinetobacter lwoffii and Pseudomonas stutzeri. If the ultrasound unit keyboard, handle, transducer, electrical cord, and connector are not regularly cleaned between patients, they may pose as a vector for transmission of potential pathogens between the operator and the patient. The entire ultrasound unit should be considered as a potential source of infection. The pooled risk of cross infection via ultrasound probes has been estimated at 3.1% of patients.

What is needed is a rapid method of assessment for cleanliness of ultrasound probes and associated equipment. Cleanliness testing using rapid adenosine triphosphate (ATP) equipment has been suggested as superior to both visual inspection and microbial sampling because it provides a real-time and quantitative measure of cleanliness. Although the use of ATP testing is becoming more common for assessing instrument cleanliness, there remains concerns over applicability, precision, and variability. Unfortunately, many articles written around ATP testing do not account for inherent variability, and propose evidence that is consequently unsustainable.

To overcome the difficulties with variability, a new algorithm-based sampling method has been proposed to mitigate the problems encountered in field use with ATP testing. This method requires multiple samples and also includes a cleaning step to internally validate the cleanliness and cleanability of the surface or device being assessed. This limited scope study sought to classify the devices and surfaces tested into 3 broad groups: clean, equivocal (probably unclean), and dirty (definitely unclean).

Participants in the study were contacted through the ASUM with the clearly stated outcome of anonymity and peer review publication of the results. This article outlines this research project in anticipation of further and more detailed follow-up studies into this important area of infection prevention and iatrogenic risk.

METHODS

ATP testing was conducted at 5 hospital ultrasound clinics within Sydney, Australia. Within each clinic, individual ultrasound suites were selected on the basis of availability after cleaning and high-level disinfection where appropriate. At some locations, the ultrasound instruments and probes were stored in an adjacent room. In each situation, the equipment tested was confirmed by staff as patient ready for use.

Measurement method

ATP testing was conducted on 3 separate days using ATP bioluminometer and associated swabs (Hygiena; Key Diagnostics, Sylvania, Australia). ATP testing devices express results in a relative light unit (RLU) scale. The Hygiena device was selected after validation experiments confirmed that the point of a zero reading for ATP (0 RLU) equates to a repeatable outcome in terms of quantitated ATP measurements. The reproducible precision at the lower level of the dynamic range was important when distinguishing clean surfaces from less clean surfaces at the lower limit of quantitation for the ATP testing device.

An initial cleanliness threshold, which is specific to the Hygiena ATP testing device, was set at 100 RLUs. The level of 100 RLUs has been recommended by others, despite differences in sampling areas used of 100 and 10 cm².

The dimensions of the swabbing areas recommended for ATP testing have varied from a 100 cm² area (10 × 10 cm) in both food and health care, whereas other authors have chosen a smaller area of 16 cm² (4 × 4 cm). Using a swab area of 2 × 5 cm (10 cm²) has also been recommended for both food and health care surfaces. The 10 cm² rationale is practical for health care surfaces and reusable medical devices such as ultrasound probes and allied equipment.

Swabbing method

Using an aseptic technique, a fresh swab was uncapped and the distal tip was applied in a rolling action across a 10-cm² sampling area. The swab was then recapped, the reagent was released and mixed for 5-10 seconds, and the swab was placed into the bioluminometer, and the detection system was activated. The readings were available after 15 seconds and recorded both manually and stored within the Hygiena ATP device memory.

Stage 1: ATP testing

Our previous research has concluded that repeated testing as outlined in a sampling algorithm is required to mitigate sampling and inherent error.

In this study, each of the selected surfaces was sampled in duplicate on adjacent segments of the surface with each sample matched for sampling area (in most instances an area of 2 × 5 cm = 10 cm²). The second ATP sample was taken on all surfaces on an adjacent area. Equivocal results arose where duplicate results indicated one reading was above the 100-RLU threshold and one was below the 100-RLU threshold. Where results were equivocal or there was visible soiling present, a third ATP swab was taken.

Stage 2: Cleaning intervention step

Our hypothesis was that surfaces classified as dirty, or equivocal, could be shown to have a residual presence of ATP soil that was readily removable. This can be demonstrated through a validated cleaning intervention step (CIS), examined using ATP on a before and after basis.

This step provides evidence on the potential for achieving a cleaner surface if the cleaning is conducted with a controlled aseptic technique. After the initial sampling (duplicate or triplicate), a CIS was conducted. The CIS was not conducted on clean surfaces where the duplicate samples were both <50 RLUs.

This step used disposable detergent wipes (neutral pH) which had been validated as suitable for use with the ATP testing swabs (Speedy Clean wipes, or Matrix Wipes; Whiteley Corporation, Sydney, Australia). Disposable detergent wipes have a validated role in surface cleaning. This wiping process is not intended to replace sanitization of the surface, but rather is used only to clean away any ATP-rich residue that might be present on the surface in the area of sampling.

The principle used when cleaning with the disposable wipe was to use only 1 wipe, on 1 surface, wiped in 1 direction. The wipe was used by first removing the wipe using aseptic technique (including hand hygiene with an alcohol-based handrub), and then one side of the wipe was rubbed broadly across the sampling area of the implement or across an area of >10 cm² for a surface to fully wet the sampling area. The wipe was then folded in half with the unused side on the outer aspect. The disposable wipe was then wiped
in a single direction taking care not to allow the wipe to contact the surface beyond the moistened area (therefore avoiding recontamination of the cleaned area). The second unused side of the disposable wipe was then used for the second wiping action on and immediately adjacent to the surface region where the first wipe was completed following Rutala et al. This second wipe removed excess liquid and allowed for faster drying of the surface to be retested.

Stage 3: ATP retest

The moistened area was allowed to air-dry before a postclean ATP sample was taken by swabbing inside the freshly cleaned (wiped) surface area using the 2 × 5-cm swabbing pattern. The goal of the CIS was to achieve an ATP reading of not more than 50 RLUs and if possible ≤25 RLUs.

The reasoning behind this threshold arises from the random variation exhibited by ATP testing devices. The CIS was repeated and the area resampled where the post-CIS reading was >50 RLUs, and this process continued until the post-CIS reading was ≤50 RLUs.

Every surface had an initial duplicate measurement recorded in RLUs. Using duplicate sampling does not overcome typical sampling error, but it does reduce the impact of inherent variability when testing with ATP systems. Following our previously published protocol, we defined an interim outcome with 5 categories as follows:

- **Clean (tertiary):** 2 ×< 25 RLUs: A very high likelihood this surface is clean.
- **Clean (secondary):** 2 ×<50 RLUs: High likelihood that this surface is clean and not dirty.
- **Clean (primary):** 2 ×<100 RLUs: Most likely this surface is clean and not dirty.
- **Equivocal:** (1 each ±100 RLUs): Likely that this surface is dirty.
- **Dirty (2 ×>100 RLUs):** Almost certain this surface is dirty.

Surfaces with equivocal results were subjected to a third swab and redistributed as appropriate to allow an outcome with 4 final categories.

Statistical analyses

Descriptive results for categorical variables are presented as frequency counts and percentages. The only numerical outcome, ATP-level measures in RLUs, is skewed to the right. Descriptive statistics are presented as medians and quartiles with graphic results (box plots and scatterplot) and are presented with logarithmic axis scales. Statistical comparison of ATP measured in RLUs before and after cleaning is conducted using the Wilcoxon matched-pairs test.

Statistical analyses and graphs were completed initially using Microsoft Excel 2013, v7.0 (Microsoft, Redmond, WA) and SPSS version 22.0 (IBM, New York, NY). Before and after analysis was completed in Sigma Plot 13.0 (Systat, San Jose, CA).

RESULTS

The survey was conducted across 5 different sites labeled A-E. The ATP testing was conducted on 253 surfaces, including a range of patient-ready ultrasound probes. In total, >750 ATP swabs were conducted. Table 1 shows the number of surfaces tested at each site, and lists which equipment was tested. Figure 1 shows the observed results from each of the 253 RLU duplicates and how these measures classify the cleanliness of these surfaces.

Of the surfaces tested, 26% (66/253) was classified as either dirty or equivocal. Three of the surfaces with equivocal results are missing follow-up testing and therefore are missing in the final cleanliness measurement. Equivocal results (41/253) were then retested (38/250) and redistributed to one of the other categories leading to a final outcome of 4 categories of surface cleanliness as shown in Tables 2 and 3.

To confirm that surfaces really were dirty, each dirty surface was cleaned following the standard cleaning protocol and retested. For comparison, we applied the same cleaning and retesting for 151 of the surfaces, including most of the equivocal surfaces (38/41), all of the dirty surfaces (25), and a selection of the clean or very clean surfaces (85).

Table 2 shows cleanliness category of each surface by type of equipment. Counts and row percentages are provided. Notice that all equipment was found to be dirty more than once. After discounting the results for screen because of small sample sizes, the gel bottle (46%) and chairs (44%) were most likely to show dirty surfaces.

Table 3 shows cleanliness category of each surface by participating site. Dirty surfaces were observed across all sites.

There were 148 surfaces subjected to cleaning and retesting with an ATP test. The results of testing after the first cleaning are summarized in Figure 2. Notice that except for a single outlier in the primary clean group, all surfaces recorded results <100–RLU cleanliness threshold after a single cleaning episode, with most ≤25 RLUs. The median RLUs in the dirty group after cleaning (4 RLUs) was essentially identical to the medians of all 3 clean groups (5, 4, and 5 RLUs). The fact that cleaning addresses the problem, validates the fact that the dirty surfaces were dirty and required cleaning.

Details of the 148 surfaces with results before and after cleaning are shown in Table 4. The results indicated that for the 13 items where the initial results from duplicate swabbing were below an average of 25 RLU, the CIS did not lower ATP any further (P = .869) (Wilcoxon). For all of the other surfaces, the CIS significantly reduced ATP readings (P < .001).

Of the 148 items and surfaces tested, 9 (6%) required ≥2 repeat CISs to reduce the ATP level to below the desired 25 RLUs. For 3 surfaces, the ATP testing result increased after the cleaning step. The results from these items are presented in Table 5.

DISCUSSION

The physical appearance of a clean and patient-ready reusable medical device is not a reliable indicator of actual cleanliness. The use of ATP testing as a quality assurance indicator has been demonstrated to be superior to visual inspection, and the results from this study corroborate this observation. The novel ATP sampling algorithm used in this study minimized the impact of inherent variability arising from the ATP testing devices. The CIS allowed an
Table 2
Cleanliness category by type of equipment

<table>
<thead>
<tr>
<th>Equipment</th>
<th>Clean (tertiary)</th>
<th>Clean (secondary)</th>
<th>Clean (primary)</th>
<th>Dirty</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe head</td>
<td>46 (56.1)</td>
<td>12 (14.6)</td>
<td>15 (18.3)</td>
<td>9 (11.0)</td>
<td>82 (100)</td>
</tr>
<tr>
<td>Probe connector</td>
<td>13 (29.5)</td>
<td>16 (36.4)</td>
<td>13 (29.5)</td>
<td>2 (4.5)</td>
<td>44 (100)</td>
</tr>
<tr>
<td>Chair</td>
<td>11 (39.3)</td>
<td>2 (71)</td>
<td>3 (10.7)</td>
<td>12 (42.9)</td>
<td>28 (100)</td>
</tr>
<tr>
<td>Machine grip</td>
<td>5 (20.8)</td>
<td>6 (25.0)</td>
<td>7 (29.2)</td>
<td>6 (25.0)</td>
<td>24 (100)</td>
</tr>
<tr>
<td>Keyboard</td>
<td>12 (29.3)</td>
<td>15 (36.6)</td>
<td>10 (24.4)</td>
<td>4 (9.8)</td>
<td>41 (100)</td>
</tr>
<tr>
<td>Gel bottle</td>
<td>1 (3.6)</td>
<td>5 (17.9)</td>
<td>9 (32.1)</td>
<td>13 (46.4)</td>
<td>28 (100)</td>
</tr>
<tr>
<td>Screen</td>
<td>1 (33.3)</td>
<td>0 (0)</td>
<td>2 (66.7)</td>
<td>3 (100)</td>
<td>3 (100)</td>
</tr>
<tr>
<td>Total</td>
<td>89 (35.6)</td>
<td>56 (22.4)</td>
<td>57 (22.8)</td>
<td>48 (19.2)</td>
<td>250 (100)</td>
</tr>
</tbody>
</table>

NOTE. Values are n (%).

Table 3
Cleanliness category by participating site

<table>
<thead>
<tr>
<th>Hospital</th>
<th>Clean (tertiary)</th>
<th>Clean (secondary)</th>
<th>Clean (primary)</th>
<th>Dirty</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>39 (52.7)</td>
<td>14 (18.9)</td>
<td>15 (20.3)</td>
<td>6 (8.1)</td>
<td>74 (100)</td>
</tr>
<tr>
<td>B</td>
<td>10 (26.3)</td>
<td>9 (23.7)</td>
<td>7 (18.4)</td>
<td>12 (31.6)</td>
<td>38 (100)</td>
</tr>
<tr>
<td>C</td>
<td>25 (29.8)</td>
<td>22 (26.2)</td>
<td>15 (17.9)</td>
<td>84 (100)</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>9 (24.3)</td>
<td>4 (10.8)</td>
<td>14 (37.8)</td>
<td>37 (100)</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>6 (35.3)</td>
<td>7 (41.2)</td>
<td>3 (17.6)</td>
<td>17 (100)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>89 (35.6)</td>
<td>56 (22.4)</td>
<td>57 (22.8)</td>
<td>48 (19.2)</td>
<td>250 (100)</td>
</tr>
</tbody>
</table>

NOTE. Values are n (%).

Table 4
Cleanliness classification including median and interquartile range

<table>
<thead>
<tr>
<th>Cleanliness sampling location</th>
<th>Clean (tertiary)</th>
<th>Clean (secondary)</th>
<th>Clean (primary)</th>
<th>Dirty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average RLU prior to cleaning</td>
<td>7.5 (1.0-13.0)</td>
<td>32.5 (26.5-36.0)</td>
<td>53.5 (42.5-70.5)</td>
<td>147.8 (121.8-206.0)</td>
</tr>
<tr>
<td>First postclean swab result</td>
<td>5.0 (0-11.0)</td>
<td>4.0 (10.7-5.0)</td>
<td>5.0 (2.5-10.0)</td>
<td>4.0 (2.0-11.0)</td>
</tr>
<tr>
<td>Difference from first swab to first postcleaning swab—Wilcoxon matched-pairs test (P)</td>
<td>.208</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

NOTE. Values are median (interquartile range) or as otherwise indicated.

RLU, relative light unit.
immediate assessment of the cleanliness potential and demonstrated that cleanliness could be significantly improved (P < .001) on 91.2% (135/148) of the surfaces cleaned and retested.

All of the surfaces and devices tested were either reusable medical devices intended for intact skin (noncritical) or surfaces or objects within the field of operation for the provision of ultrasound services. A limited number of the ultrasound probe heads were semi-critical devices intended for intracavity use. The ASUM has recently introduced new guidelines intended to ensure that minimum cleanliness standards are maintained for patient-ready ultrasound devices; although these guidelines do provide an indication of cleanliness required, verifying these standards in practical and field-based situations is problematic and relies entirely on visual inspection.¹

The findings from this study indicate that cleanliness standards achieved through existing quality assurance techniques (which rely on visual inspection) are insufficient. Cleaning is a process that is intended to remove unwanted material from a surface or place.³⁴

The findings presented in this limited scope study suggest that only those surfaces classified as clean (tertiary) could not be improved by further cleaning. Only 13 of 253 (4.3%) surfaces tested exhibited this level of cleanliness. Although 74% (187/253) of surfaces tested achieved the literature-recommended 100 RLUs, >1 in 4 surfaces (66/253) were found to fail this guideline with at least 1 ATP swab result.

The results of a before and after testing approach used with the ATP sampling algorithm ensured that both inherent and sampling error were minimized, therefore increasing the certainty over the validity of the results.

One disturbing finding arising from the use of a CIS was the requirement to repeat the CIS for several of the items and surfaces tested. For 6% (9/148) of surfaces subjected to the CIS, cleaning difficulty was a problem and these surfaces required repeated CISs. This indicated that either the surface was difficult to clean, or in one instance (a semi-critical ultrasound probe), the surface was residually unclean and required 4 separate cleaning efforts to reduce the ATP level to below the desired 25 RLUs. This cleanliness failure may have been because of poor instrument design and cleanability, or because of cleaning failure, but this study did not further investigate this finding.

Just as alarming were the findings that in 3 instances, the ATP result went up after the CIS. This may be because of inherent variability in one of the devices, but for the keyboard in hospital D, and the semi-critical intracavity device in hospital C, the readings continued to remain high despite repeated cleaning steps. This most likely indicates the liberation of additional soils deposited on those surfaces that are not visible to the naked eye.

The implications of this work for a busy ultrasound department are an important consideration for ultrasound staff and
infection preventionists. The cost and workload efficiencies of a busy ultrasound department will require a carefully considered sampling plan as part of a quality assurance program. Clearly not every piece of equipment can be sampled, but the frequency of sampling, the repeated sampling of typical indicators, and randomized sampling among other potentially contaminated surfaces and objects within the patient treatment zone can be instituted economically to suit the risk profile of the individual institution and provide an accurate assessment of the overall efficiency of the cleaning and decontamination processes within the department. This work also has wider implications for other reusable medical devices that require cleanliness assessments as part of infection prevention.

The strengths of this study are that it was conducted using scientifically validated protocol with all measurements and interventions by the same duo who are highly experienced in the use of these devices. The limitations of this study include the small number of possible health care sites within the same urban area. The work would also have been enhanced with a parallel microbiological sampling program. Further work is recommended to improve and validate the cleanliness standards with a broader study including presence of other soiling materials.

CONCLUSIONS

This study investigated the cleanliness standards of patient-ready ultrasound equipment and found that even a simple cleaning process would improve cleanliness standards in 96% of patient-ready-for-reuse ultrasound equipment. ATP testing using the algorithm approach improved the certainty around the results and mitigated both the inherent error of the devices and the sampling error that would occur through use of a single sampling point approach. The work suggests further studies are required to establish suitable benchmarks for a simple quality assurance process covering reusable medical devices such as ultrasound equipment.

References