Development of an observational measure of healthcare worker hand-hygiene behaviour: the hand-hygiene observation tool (HHOT)

J. McAteer, S. Stone, C. Fuller, A. Charlett, B. Cookson, R. Slade, S. Michie, and the NOSEC/FIT group

Department of Psychology, University College London, London, UK
Royal Free & University College Medical School, London, UK
Department of Primary Care and Population Science, University College London, London, UK
Statistics, Modelling & Bioinformatics Department, Health Protection Agency, London, UK
Laboratory of Healthcare Associated Infection, Health Protection Agency, London, UK

Received 25 September 2007; accepted 14 December 2007
Available online 4 March 2008

KEYWORDS
Hand-hygiene behaviour; Healthcare-associated infection; Observational measurement

Summary
Previous observational measures of healthcare worker (HCW) hand-hygiene behaviour (HHB) fail to provide adequate standard operating procedures (SOPs), accounts of inter-rater agreement testing or evidence of sensitivity to change. This study reports the development of an observational tool in a way that addresses these deficiencies. Observational categories were developed systematically, guided by a clinical guideline, previous measures and pilot hand-hygiene behaviour observations (HHOs). The measure, a simpler version of the Geneva tool, consists of HHOs (before and after low-risk, high-risk or unobserved contact), HHBs (soap, alcohol hand rub, no action, unknown), and type of HCW. Inter-observer agreement for each category was assessed by observation of 298 HHOs and HHBs by two independent observers on acute elderly and intensive care units. Raw agreement (%) and Kappa were 77% and 0.68 for HHB; 83% and 0.77 for HHO; and 90% and 0.77 for HCW. Inter-observer agreement for overall compliance of a group of HCWs was assessed by observation of 1191 HHOs and HHBs by two independent observers on acute elderly and intensive care units. Raw agreement (%) and Kappa were 77% and 0.68 for HHB; 83% and 0.77 for HHO; and 90% and 0.77 for HCW. Inter-observer agreement for overall compliance of a group of HCWs was assessed by observation of 1191 HHOs and HHBs by two independent observers. Overall agreement was good (intraclass correlation coefficient = 0.79). Sensitivity to change was examined by autoregressive time-series modelling of...
longitudinal observations for 8 months on the intensive therapy unit during an *Acinetobacter baumannii* outbreak and subsequent strengthening of infection control measures. Sensitivity to change was demonstrated by a rise in compliance from 80 to 98% with an odds ratio of increased compliance of 7.00 (95% confidence interval: 4.02–12.2) $P < 0.001$.

© 2008 The Hospital Infection Society. Published by Elsevier Ltd. All rights reserved.

Introduction

Levels of healthcare-associated infection are high in the UK,\(^1\,^2\) Improved hand-hygiene behaviour (HHB) reduces infection and hospitals have a legal duty to ensure good hand-hygiene compliance.\(^3\,^4\,^5\) Audit of compliance embedded in clinical governance through the 'Saving Lives' campaign and the national CleanYourHands campaign encourages healthcare workers (HCWs) to increase HHB.\(^6\,^7\) In addition, a national trial is evaluating an intervention to increase HHB.\(^8\) A valid, reliable, simple and sensitive measure of HHB is therefore needed for audit, service improvement and research.

Direct observation is considered the gold standard of HHB measurement but recent systematic reviews have questioned the methodological quality of available measures, particularly the lack of information on inter-rater reliability and clear standard operating procedures which are required to assess reliability and ensure widespread replication.\(^9\,^10\) For example, a definition of one hand-hygiene behaviour observation (HHO) as ‘activities involving indirect patient contact or hospital maintenance’ is not sufficiently specific to ensure agreement between observers as to whether, 'rearranging objects on a patient’s bedside table, without directly touching the patient’, is in this category or not.\(^11\)

For studies testing inter-rater reliability, procedures used to ensure observer coding of identical single HHOs and HHBs are unclear.\(^11,^12\) For example, observer A may observe and record one hand-hygiene opportunity and subsequent HHB, whereas observer B may have missed that particular opportunity, but recorded another patient contact and subsequent HHB. According to the observation sheets of observer A and observer B, this would look like 100% agreement, when in fact there was none. Procedures used to assess inter-rater reliability must allow for reliable pairing of observer codings. Additionally, studies do not assess sensitivity to change.

This paper describes the development of a tool to measure HHB, its standard operating procedures (SOPs), and assessment of its inter-rater reliability and sensitivity to change.

Methods

Developing and defining observational categories and rules

Observational categories and rules for coding observed events into categories were developed by drawing on three sources of information.

Guideline for HHB

The UK Department of Health Epic project guidelines for HCW HHB formed the basis of the measure: 'Hands must be decontaminated immediately before each and every episode of direct patient contact/care and after any activity or contact that potentially results in hands becoming contaminated'.\(^13\)

Pilot observations

Pilot observations were conducted by three researchers [a research nurse (C.F.), health psychologist (J.M.) and physician (S.S.)] in one acute care of the elderly (ACE) unit and one intensive therapy unit (ITU) at two inner London NHS hospitals. These are settings with vulnerable patients, high incidence of hospital-acquired infection and frequent close patient contact, including high-risk opportunities. Ethical approval was obtained from the ethics committees of both hospitals.

Initially, the Lewisham tool and the Geneva tool were piloted.\(^11,^14\) The former records opportunities for HHB across 21 observational categories; we found observer disagreement to be common because of the large number of categories recorded. The Geneva tool, with only 13 categories, records HCWs, their HHBs and HHOs. Ten hours of observational data were collected, 5 h on each ward, 1 h at a time, noting situations posing problems for observing or categorising.
Observation was conducted within a predefined clinical area of four to six beds during two high-frequency patient contact periods in the morning and the afternoon (9:00–12:00 and 14:00–16:00).

**Measure refinement**

Observational data were discussed and categories simplified by the observers, a consultant microbiologist/infection control expert (B.D.C.) and a health psychologist (S.M.). Clearly defined categories for observed HHBs, opportunities and HCWs were developed. Problem situations were discussed, solutions agreed and definitions modified accordingly with members of the project steering group, many of whom had experience or specialised knowledge of the relevant issues. Observation and discussion continued iteratively until we were satisfied that the measure, a simpler version of the Geneva tool, was able to record all HHBs, HHOs and HCWs in a wide range of clinical situations (Figure 1), using the SOPs summarised in Table I. Full SOPs are available elsewhere (www.idrn.org//nosec.php).

**Inter-observer agreement for individual HHOs and their subsequent HHBs**

A sample size of 200 hand-hygiene opportunities was required to ensure precision of ±0.1 around the estimated Kappa statistic. This is based on the expected Kappa statistic ranging between 0.7 and 0.9, and the 'proportion positive' by each observer ranging between 0.5 and 0.75. This was based on data from the pilot observations of 95 hand-hygiene events (i.e. linked HHB, HHO and HCW) observed by two pairs of observers (47 and 48 respectively) with HHB as the primary outcome measure (Kappa statistics 0.92 and 0.91 respectively). Extrapolation on the basis that one-third of observed opportunities would be classified as 'unknown' (i.e. not clearly observed) increased the required number of patient contacts to around 300. These calculations were performed using the `sskdlg` command in Stata 8.2.

In order to test inter-observer agreement, two trained observers observed identical hand-hygiene opportunities. Six two-hour long observation sessions were conducted at one intensive care unit and one acute care of the elderly ward at one London teaching hospital. Observers sat near each other in a 4–6 bed area, but could not see the other's recordings. Each HHB, HHO and HCW occurring within a predefined clinical area of 4–6 beds on each ward was recorded. Each opportunity was recorded on a separate observation sheet. Upon completion of an opportunity — as signalled by an HHB — observers would confer as to whether or not they observed the same opportunity. If yes, observer classifications were paired with one observer stapling two observation sheets as one pair. If no, the classified hand-hygiene opportunity and

![Table 1](http://www.idrn.org/nosec.php)

Before low-risk contact | After low-risk contact | Before high-risk contact | After high-risk contact | Before unobserved contact | After unobserved contact
---|---|---|---|---|---
Doctor | | | | | |
Indication | x | | | | |
Soap | x | | | | |
Alcohol | | | | | |
No action | | | | | |
Unknown | | | | | |
Nurse/HCA | | | | | |
Indication | x | | | | |
Soap | | | | | |
Alcohol | | | | | |
No action | | | | | |
Unknown | | | | | |
Other/Unsure | | | | | |
Indication | | | | | |
Soap | | | | | |
Alcohol | | | | | |
No action | | | | | |
Unknown | | | | | |

![Figure 1](http://www.idrn.org/nosec.php)  
Hand-hygiene observation tool.
behaviour was paired — and thus stapled — with a blank observation sheet. A total of 298 observations were collected. Raw measures of agreement (%) and unweighted Kappa coefficients were calculated for HHB, HHO and HCW (<0 = less than chance agreement; 0.01–0.20 = slight agreement; 0.21–0.40 = fair agreement; 0.41–0.60 = moderate agreement; 0.61–0.80 = substantial agreement; 0.81–0.99 = almost perfect agreement).16

Inter-observer agreement for overall HHB compliance

Four trained observers (J.M., C.F., S.S., R.S.) carried out 19 h of observation (1191 HHOs) between them working in pairs observing HCWs, their HHOs and HHBs on the above wards and at the times previously described. Each pair of observers agreed the area of the ward to be observed, and made independent assessment of HCWs, HHOs and HHBs for an hour at a time, without reference to each other, recording multiple events on one sheet. Observers sat near each other to ensure a similar vantage point. Each observer’s assessment of overall compliance of HCWs observed during the hour was recorded. Analysis was by the intraclass correlation, estimated using the loneway command in Stata, after applying an arc sin root transformation of the proportion compliant.
Sensitivity to change

A trained observer (S.S.) observed hand-hygiene compliance on one ITU for one hour during the morning each month for nine months (October–June inclusive) as part of a pilot study to develop an intervention to improve hand hygiene. During February a prolonged outbreak of Acinetobacter baumannii resulted in a major emphasis on infection control with measures such as demarcation of the clinical area around each bed by red tape on the ground, within which area all staff and visitors had to wear gown and gloves and increase emphasis on hand-hygiene. Analysis was by a generalised estimating equation model, which assumed an AR1 autoregressive correlation structure with the outcome variable being compliance. The predictor variables were an indicator variable for post outbreak, and month as a continuous variable from 1 to 8, ignoring the fact that there is no observation in May. The interaction between post outbreak and time was assessed. This model was implemented using the *xtgee* command in Stata.

Results

Inter-observer agreement for individual HHOs and their subsequent HHBs

Raw agreement (%), estimated Kappa and 95% confidence intervals (CI) were 76% and 0.68 (95% CI: 0.61–0.74) for the HHB category (Table II), 83% and 0.77 (95% CI: 0.71–0.83) for the opportunity category (Table III), and 90% and 0.77 (95% CI: 0.69–0.85) for the HCW category (Table IV). This indicated substantial agreement for all categories. Inspection of the raw data indicated the nature of disagreements within the HHB category. Most disagreements concerned hand-hygiene events in which HHB was assessed as ‘no action’ and ‘unknown’. This was confirmed in a secondary analysis conducted using only those hand-hygiene events in which HHB was clearly seen or not seen (i.e. excluding ‘unknown’) (N = 202). Inter-observer agreement and Kappa coefficient were 92% and 0.88 (95% CI: 0.78–0.98) for the HHB category.

Inter-observer agreement for overall HHB compliance

The mean absolute difference in overall compliance as assessed by individual observers was 12% (range: 1–40%). Overall agreement was very good with an intraclass correlation for the 19 pairs of 0.79 (95% CI: 0.62–0.96). There was an inverse relationship between the number of HHBs observed (i.e. with ‘unknowns’ excluded) and the size of the differences in observed compliance, with differences of ≥10% most likely to be seen if there were <15 HHBs observed per hour (Figure 2).

Sensitivity to change

Assessment of interaction between post outbreak and time showed no evidence that the temporal trend differed in the pre- and post-outbreak periods (P = 0.8). A main effects model was therefore fitted which estimated that the odds of compliance increased sevenfold after, compared to before, the outbreak [7.00 (95% CI: 4.02–12.2)]. There is some weak evidence of an increasing temporal trend over the time period, with an estimated increase of 9% in the odds of compliance each month (1.09, 95% CI: 0.99–1.21; P = 0.08). There is therefore clear evidence that the HHOT is capable of detecting the increase in hand hygiene compliance that accompanied the outbreak (Figure 3).

<table>
<thead>
<tr>
<th>Table II Inter-observer agreement for hand-hygiene behaviours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer 2</td>
</tr>
<tr>
<td>Soap</td>
</tr>
<tr>
<td>Observer 1</td>
</tr>
<tr>
<td>Soap</td>
</tr>
<tr>
<td>Alcohol</td>
</tr>
<tr>
<td>No action</td>
</tr>
<tr>
<td>Unknown</td>
</tr>
<tr>
<td>No opportunity</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

* Instances in which one observer coded an opportunity but the other did not.
Discussion

This study reports for the first time the development of a robustly validated, standardised observational tool to measure HHB, with clear SOPs that will facilitate its use by others and with good evidence of its inter-rater reliability and sensitivity to change.

The strength of the study is its use of a rigorous and transparent methodology. SOPs are described clearly and are available to the infection control community through the CleanYourHandsCampaign website. Inter-rater agreement for individual HHOs and their subsequent HHBs was tested by pairing observed events, using one observation sheet per hand-hygiene event. This alters the task conditions for which the measure is intended, where one observation sheet records multiple events, which could theoretically reduce the generalisability of the findings. It is also possible that inter-rater agreement was inflated because observers conferred as to whether they had recorded the same event to facilitate pairing. However, without resource-intensive video recordings of HCW behaviour, which would have ethical problems, these issues are an inevitable consequence of the need to use a paired procedure to test inter-observer reliability accurately. To the authors’ knowledge, there is no published research addressing this issue.

No other studies appear to have recorded the substantial number of patient contacts occurring behind closed curtains (19% of our sample). Discussions with other infection control experts on the project and elsewhere have identified it to be a common issue in many healthcare systems so we decided it was important to record this variable. Although it is not possible to observe an opportunity occurring behind closed curtains for ethical reasons, it is reasonable to infer that an opportunity is taking place. Additionally, the study described the sometime failure of observers to record HHOs or HHBs, described and justified its clinical setting and was specific about the timing

<table>
<thead>
<tr>
<th>Table III</th>
<th>Inter-observer agreement for hand-hygiene behaviour opportunities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before low risk</td>
</tr>
<tr>
<td>Observer 1</td>
<td></td>
</tr>
<tr>
<td>Before low risk</td>
<td>24</td>
</tr>
<tr>
<td>After low risk</td>
<td>2</td>
</tr>
<tr>
<td>Before high risk</td>
<td>1</td>
</tr>
<tr>
<td>After high risk</td>
<td>0</td>
</tr>
<tr>
<td>Before unobserved</td>
<td>6</td>
</tr>
<tr>
<td>After unobserved</td>
<td>0</td>
</tr>
<tr>
<td>No opportunity</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>34</td>
</tr>
</tbody>
</table>

* Instances in which one observer coded an opportunity but the other did not.

<table>
<thead>
<tr>
<th>Table IV</th>
<th>Inter-observer agreement for healthcare workers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Doctor</td>
</tr>
<tr>
<td>Observer 1</td>
<td></td>
</tr>
<tr>
<td>Doctor</td>
<td>30</td>
</tr>
<tr>
<td>Nurse/HCA</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
</tr>
<tr>
<td>No opportunity</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
</tr>
</tbody>
</table>

HCA, healthcare assistant.

* Instances in which one observer coded an opportunity but the other did not.
and length of observation periods, details frequently omitted in descriptions of observational assessments of hand-hygiene compliance. Further to this, clear evidence that the measure is sensitive to change was provided. We are not aware of any other study demonstrating this. The greatest number of inter-observer disagreements for individual HHBs was between the 'no action' and 'unknown' categorisations (6.3%) of the sample. This may have occurred for two reasons. First, duration between observed opportunity and subsequent HHB varied. Observers

![Figure 2](image-url)  
**Figure 2** Absolute difference between pairs of observers vs the average compliance of the two observers when estimating overall compliance of a group of healthcare workers.

![Figure 3](image-url)  
**Figure 3** Regression model estimate of the proportion of opportunities in which healthcare workers complied with hand-hygiene. Hollow circles: observed; filled circles: model estimate; bars: 95% confidence limits.
tracked multiple HCWs until they engaged (or not) in HHB, moved to another patient contact, or exited the field of observation. For example, observer A and B tracked opportunities 1–5. Observer A coded opportunity 1 as no action. Observer B lost track of opportunity 1, whilst tracking 2–5, and thus coded HHB for opportunity 1 as 'unknown'. Second, observers reported difficulty in detecting personal hand-rub dispensers, i.e. those attached to the HCW’s clothing. For example, observers 1 and 2 tracked an unobserved opportunity, i.e. occurring behind the curtain. Observer 1 noticed that the HCW was wearing a personal hand-rub dispenser, but Observer 2 did not. The HCW re-emerged and exited the field of observation. Observer 1 coded HHB as 'unknown', whereas observer 2 coded this as 'no action' having not observed the dispenser. Personal dispensers may therefore increase the difficulty of monitoring compliance.

This study did not, in common with most observational hand-hygiene measures, assess quality of hand-hygiene or describe the training required to use the tool. Nonetheless it is worth commenting that subsequent experience has indicated that training takes between 4 and 6 h. In assessing the reliability of overall compliance, the four instances of >20% difference between observers may have been due to training effects. The time required to train an observer needs further formal study, but is beyond the scope of this paper.

In conclusion, this study has addressed three issues: the need for clear and replicable SOPs, transparent testing of inter-observer agreement, and demonstration of sensitivity to change. The resultant HHOT, a simplified version of the Geneva tool, is suitable for behavioural interventions, individual and group audit, service improvement and research.

Acknowledgements

We thank staff at the acute care of the elderly unit and intensive care unit at the Royal Free Hospital, London; D. Pittet, University of Geneva Hospitals, Switzerland for providing a copy of the Geneva tool; and R. Sladek, Australian Centre for Evidence Based Clinical Practice for commentary on earlier drafts of the manuscript.

Conflict of interest statement

None declared.

Funding sources

Funded by the Hospital Infection Society, Patient Safety Research Programme, and the Royal Free Trustees. An unrestricted educational grant was provided by Gojo Industries.

References