Effect of Temperature and Time on Fecal Hemoglobin Stability in 5 Fecal Immunochemical Test Methods and One Guaiac Method

Peter Catomeris, PhD; Nancy N. Baxter, MD, PhD; Sheila C. Boss, PhD; Lawrence F. Paszat, MD, PhD; Linda Rabeneck, MD, MPH; Edward Randell, PhD; Mardie L. Serenity, MHSc, PMP; Rinku Sutradhar, PhD; Jill Tinmouth, MD, PhD

Context.—Although promising for colorectal cancer screening, hemoglobin (Hb) stability remains a concern with fecal immunochemical tests. This study implemented a novel, standardized method to compare Hb stability across various fecal immunochemical tests. The method can be used to inform decisions when selecting a kit for use in colorectal cancer screening. In so doing, this work addressed a critical need for standardization in this field.

Objective.—To compare the stability of Hb across 5 different immunochemical kits and one guaiac kit.

Design.—The stability of Hb was analyzed in collection devices inoculated with Hb-spiked feces and (1) stored at different immunochemical kits and one guaiac kit. The method was repeated by courier or postal services in uncontrolled temperature conditions from 3 locations in Ontario, Canada, to a central testing center.

Results.—The stability of Hb varied with time and temperature and by kit. Lower Hb recoveries occurred with increasing temperature and increasing time from sample collection to testing. Refrigeration provided the best stability, although results varied across kits (eg, from 4.2 days to >60 days before a prespecified threshold [<70% probability of the test results remaining positive] was reached). Freeze-thaw stability varied across kits and cycles (Hb recoveries: NS Plus [Alfresa Pharma, Chuo-ku, Osaka, Japan], 91.7% to 95.4%; OC Diana [Eiken Chemical, Taito-ku, Tokyo, Japan], 57.6% to 74.9%). Agreement regarding Hb levels before and after transportation varied across kits (from 57% to 100%).

Conclusions.—Important differences in Hb stability were found across the included fecal immunochemical tests. These findings should inform practice-based and population-based colorectal cancer screening.

Screening with guaiac-based fecal occult blood tests (gFOBT), followed by colonoscopy for patients who test positive, was associated with a reduction of 15% to 33% in colorectal cancer (CRC)–related mortality in 3 large, randomized clinical trials. Fecal testing for occult blood is the most commonly used screening method in CRC screening programs worldwide. Until recently, those programs predominantly used kits based on the guaiac reagent to detect the heme moiety of blood, relying on the inherent peroxidase activity of heme. Unfortunately, there are important drawbacks to the guaiac method, including low analytic sensitivity, low analytic specificity, and low clinical sensitivity.

Promising fecal immunochemical test (FIT) methods have been developed. These immunoassays form an antibody-antigen complex with the globin moiety of human hemoglobin (Hb) to generate either an immunoturbidimetric signal, which can be measured (quantitative) or by lateral-flow immunochromatographic analysis, a band that can be detected visually (qualitative). The FITs have been shown to have better analytic sensitivity and specificity than gFOBT, as well as better clinical sensitivity, with minimal loss of clinical specificity. In addition, participation...
with screening is improved with FIT compared with gFOBT, likely because fewer samples (1 instead of 3) are required, and there are no dietary restrictions. Human Hb continues to degrade after feces are passed and delay of sample testing can cause false-negative results in both gFOBT and FIT methods. FITs appear to be particularly unstable, as reported recently in an Italian CRC screening program; furthermore, the wet collection methods used by some FIT kits may exhibit more Hb degradation than dry collection methods.

Organized CRC screening programs require Hb to remain stable within the collection devices for some time because the sample is typically collected at home and returned, often by mail, to the laboratory for testing. The FIT manufacturers’ recommendations, as well as data from studies of FIT stability, suggest shorter intervals between collection and testing might be needed if FIT is used instead of gFOBT. However, variation in Hb stability has been reported across kits. It is possible that acceptable times between sample collection and testing could also vary by FIT brand.

To date, there are limited data comparing the Hb stability in a standardized fashion for various quantitative and qualitative FITs, making informed selection of a FIT for use in organized CRC screening programs difficult. The aim of the current study was to examine in a standardized manner the stability of human fecal Hb in the collection devices of 5 FIT kits and 1 gFOBT kit over time and at several different temperatures.

**Materials and Methods**

**Terminology**

The following terminology is used throughout. Kit refers to the package of components, including collection devices, required for the collection and analysis of Hb in fecal samples. Collection device is the component of a kit to which the feces are added and stored until analysis. Spiked feces is a fecal sample to which Hb has been added in the laboratory. Inoculated device is a collection device to which feces (spiked or nonspiked) has been added for analysis of Hb.

**Overview**

Three in vitro studies were conducted from February to April 2012 to assess Hb stability. Study 1 was a controlled study of inoculated devices stored at various controlled temperatures (frozen, refrigerated, ambient, and elevated) and over time up to 60 days. Study 2 was a controlled study of inoculated devices exposed to several freeze-thaw cycles. Study 3 was a field study in which inoculated devices were transported by courier or postal services under uncontrolled temperature conditions, which was intended to simulate conditions that inoculated devices would experience in the CRC screening program in Ontario, Canada.

**Approval**

This study was reviewed and approved by the Sunnybrook Research Institute (Toronto, Ontario, Canada) research ethics board.

**Kits**

Table 1 summarizes the characteristics of the kits included in the current study. The FIT kits were selected for inclusion if they were approved for use by Health Canada and if they were suitable for use in an organized, population-based CRC program (ie, they were point-of-care kits). The following FIT kits were included: (1) Hb NS-Plus, manufactured by Al fresa Pharma Corporation (Chuo-ku, Osaka, Japan); (2) OC-Sensor Diana, manufactured by Eiken Chemical Co, Ltd (Taito-ku, Tokyo, Japan); (3) Hema-Screen SPECIFIC, manufactured by Immunostics, Inc (Eatontown, New Jersey); (4) FOB Advanced, manufactured by ulti med (A hrensburg, Germany); (5) Hemoccult ICT, manufactured by Beckman Coulter, Inc (Brea, California); and (6) NS-Plus, manufactured by Alfresa Pharma Corporation (Chuo-ku, Osaka, Japan). The gFOBT (Hema-Screen Guaiac, Immunostics) currently used in the Ontario, Canada, CRC screening program was also included.

**Preparation of Spiked Feces**

For all studies, fresh human feces, without preservatives and from multiple donors, were used. Feces were spiked with human Hb from ethylenediamine tetra-acetic acid whole blood specimens. Initial Hb concentrations were determined with the Sysmex XE-2100 series automated hematology analyzers (Sysmex Corporation, Kobe, Hyogo Prefecture, Japan) using the sodium lauryl sulfate method, calibrated to International Council for Standardization in Hematology/World Health Organization international hemoglobin-cyanide standard (70600). Hemolysates were prepared by diluting whole blood 10-fold with deionized water. The Hb was transferred to the fecal material, and a homogeneous mixture was obtained by stirring the feces with wooden sticks. Collection devices were inoculated with feces at room temperature within 30 minutes of spiking with Hb. Using the procedure described,
acceptable and reproducible recoveries of Hb were obtained. When measured with the OC Diana kit, the range of Hb recovery was 85.4% to 112.7% in 16 replicates. Mean (SD) recovery was 103.4% (9.7%).

**Definitions of Stability**

To date, there is no universal or standardized approach to assess Hb stability.\(^1\)\(^2\)\(^6\) Although some manufacturers do provide estimates of Hb stability in their product inserts, those estimates cannot be reliably compared across brands because the methods to determine stability were not described in the package inserts and time frames varied. For the purposes of our analysis, we measured Hb stability using separate, but standardized, approaches for the comparison of quantitative kits alone and for comparison of all kits included in the study. These 2 approaches are described below.

For comparison of Hb stability in the collection devices of the quantitative kits only, the duration of stability in study 1 was defined as the time until the percentage of Hb recovery was reduced (from 100%) by less than 2 times the between-run precision, as determined from internal studies. The between-run precision was approximately 13% and 9% at concentrations of 33 and 100 μg Hb/g feces, respectively. Therefore, mean recovery of 80% or more was used as a cutoff for stability when comparing the 2 quantitative kits.

Because qualitative FIT and gFOBT test results are read in a binary fashion, the above approach could not be used. To compare Hb stability across all kits (both qualitative and quantitative), we defined the duration of stability as the time that elapsed until the probability of the result remaining positive dropped below a prespecified threshold. For this analysis, quantitative kit results were categorized as having positive or negative results relative to the manufacturer’s suggested positivity threshold (20 μg Hb/g feces for both kits). Stability at 3 prespecified thresholds (80% probability of remaining positive, 70%, and 50%) was measured. The clinical implications of these thresholds are not known. However, the only other study\(^17\) to establish a threshold for stability used 50%, so our thresholds match or exceed those used by others.

For all 3 studies, results were defined as positive if the Hb concentration was above the manufacturer’s recommended positivity threshold (see Table 1) for quantitative kits or if the visual band was detected for qualitative kits.

**Study 1: Controlled Sample Stability Study**

Preliminary in-house studies showed that the analytic sensitivities of the kits were 20 μg Hb/g of stool (NS Plus and OC Diana), 50 μg Hb/g of stool (Immunostics FIT), 100 μg Hb/g of stool (ulti med), 500 μg Hb/g of stool (NS Plus and OC Diana), and 1000 μg Hb/g of stool (Immunostics Guaiac). Feces were spiked to produce Hb concentrations from 25 to 3000 μg Hb/g of feces. As the analytic sensitivities (which reflect the lower limit of Hb detection) varied across kits, concentration ranges appropriate to the analytic sensitivity of each kit were selected.\(^16\)\(^17\) Final Hb concentrations of 25, 50, 100, and 200 μg Hb/g of feces were used for the NS Plus, OC Diana, and Immunostics FIT kits; 100, 200, 500, 1125, and 1500 μg Hb/g of feces were used for the ulti med kit; and 500, 1125, and 1500 μg Hb/g of feces were used for the Beckman ICT and Immunostics Guaiac kits.

Multiple inoculated devices were prepared for each kit (33 sets per kit) using feces spiked with each of the Hb concentrations specified above for that particular kit. The first set was analyzed immediately at day 0 (baseline). The other 32 sets were stored under the following temperature conditions: frozen, −20° to −15°C (−4° to 5°F) (8 sets); refrigerated, 2°C to 8°C (36°F to 46°F) (8 sets); ambient, 20°C to 22°C (68°F to 72°F) (8 sets); and elevated, 45°C (113°F) (8 sets). For each temperature condition, a set was removed and analyzed at 4, 7, 10, 14, 21, 28, 45, and 60 days after baseline. Each inoculated device was tested only once during this study.

We calculated the median percentage of Hb recovery at each time point in each of the temperature conditions for each of the 2 quantitative kits. We then examined Hb stability in 2 ways: comparison of the inoculated devices from the quantitative kits only, and comparison of the inoculated devices from all kits, both quantitative and qualitative.

For the comparison of Hb stability in collection devices of the quantitative kits only, the percentage of Hb recovery compared with the baseline were first calculated for each temperature condition and at each time point. As noted above, recovery of 80% or more was used as a cutoff for stability. Under each of the 4 temperature conditions, the relationship between the percentage of Hb recovery and time was first modeled using a linear regression model. Quadratic and cubic terms were included in the regression if they improved the model fit based on likelihood ratio tests.\(^21\) A generalized estimating equation approach was used to account for any repeated measurements taken on the same sample.\(^22\) Once the estimates of the regression coefficients were obtained, those values were then used to predict the time in days and the 95% CI when the probability of results remaining positive dropped below 80%. These analyses were repeated for the other 2 probability thresholds, specifically, the time at which the probability of results remaining positive dropped below 70% and below 50%.

**Study 2: Controlled Freeze-Thaw Study**

In real-world screening programs, patient samples may be exposed to variable temperatures and conditions, including freezing and thawing, during the transport of the samples from the patient to the testing laboratory. This study was designed to determine the effect of freezing and thawing on samples in the collection devices. The stability of Hb during freezing and thawing was studied in the NS Plus, OC Diana, and ulti med collection devices. (Kits with single-use collection devices could not be used for this study.) Fecal samples were spiked to produce samples with Hb concentrations ranging from 25 to 1200 μg Hb/g of feces, in keeping with the ranges for these 3 kits used in study 1. Inoculated devices were prepared with samples at Hb concentrations appropriate to the analytic sensitivity of each type of kit, as described in study 1. Inoculated devices (NS Plus, n = 12; OC Diana, n = 12; ulti med, n = 12) were analyzed immediately (baseline), and the same devices were then analyzed after each of 3 freeze-thaw cycles. Frozen storage conditions were −20°C to −15°C (−4°F to 5°F). Samples were thawed and refrozen after analysis at 4, 7, and 10 days after the initial storage in the freezer. Inoculated devices were kept frozen for 4, 3, and 3 days between thawing, testing, and refreezing for cycles 1, 2, and 3, respectively. The number of cycles and duration were selected because they were likely to reflect anticipated real-world conditions.

For the analysis, the median percentage of Hb recovered compared with baseline values was calculated for NS Plus and OC Diana after each of the freeze-thaw cycles. For ulti med, the percentage of inoculated devices remaining positive after each cycle was calculated.

**Study 3: Field Study**

A total of 30 fecal samples were freshly spiked with Hb at the testing location to achieve final Hb concentrations of 0 μg Hb/g of feces (n = 6), 30 μg Hb/g of feces (n = 3), 60 μg Hb/g of feces (n = 6), 125 μg Hb/g of feces (n = 3), 500 μg Hb/g of feces (n = 6), and 2000...
μg Hb/g of feces \((n = 6)\). Two batches of inoculated devices using those samples were prepared for each of the 6 kits. One batch of inoculated devices (set A, 30 kits for 180 inoculated devices) was analyzed at the testing location to obtain baseline values. The second batch of inoculated devices (set B, \(n = 180\)) was packaged for transportation to 3 distribution centers (Thunder Bay, Ottawa, and London; all in Ontario, Canada). The packages were prepared without any temperature control, that is, no cooling packs, but simply with cushioning material to prevent excessive contact within the package. The study was arbitrarily designed to have 10 packages for each center. Each package contained 6 inoculated devices (1 for each of the 6 kits), for a total of 60 devices. Each center received the same distribution of Hb concentrations.

For each center, thermometers were added to 9 of the 10 packages to determine minimum and maximum temperatures during transport and a continuous temperature data logger was placed in the 10th package. The batches of packages were delivered from the testing location to the 3 distribution centers by courier. From each of the distribution centers, the 10 packages were individually mailed back to the testing location, via Canada Post regular mail.

Upon receipt at the testing location, the temperature monitors were retrieved, the transportation temperatures were recorded, and the inoculated devices were analyzed. This study was performed in February and March 2012.

Two analyses were performed. In the first, the percentage of kits in which the result after transportation agreed with the result from the package. The study was arbitrarily designed to have 10 packages for each center. Each package contained 6 inoculated devices (1 for each of the 6 kits), for a total of 60 devices. Each center received the same distribution of Hb concentrations.

For each center, thermometers were added to 9 of the 10 packages to determine minimum and maximum temperatures during transport and a continuous temperature data logger was placed in the 10th package. The batches of packages were delivered from the testing location to the 3 distribution centers by courier. From each of the distribution centers, the 10 packages were individually mailed back to the testing location, via Canada Post regular mail.

Upon receipt at the testing location, the temperature monitors were retrieved, the transportation temperatures were recorded, and the inoculated devices were analyzed. This study was performed in February and March 2012.

Two analyses were performed. In the first, the percentage of kits in which the results of the before-transportation and after-transportation tests (set B) agreed with the expected result (based on the amount of Hb spiked into each of the initial 30 fecal samples, set A) was calculated. In the second, the percentage of kits in which the result after transportation agreed with the result from before transportation (same kits, from set B) was calculated.

### RESULTS

#### Study 1: Controlled Sample Stability Study

The median percentage of Hb recoveries in the collection devices of the 2 quantitative kits over time and at varying temperatures are shown in the Figure, a and b. Inoculated devices stored under refrigerated conditions had the greatest sample stability for both NS Plus and OC Diana. Freezing the devices produced an initial decrease of 20% in Hb recovery, but then Hb recovery remained stable over time thereafter for that temperature condition. There were decreasing recoveries with increasing temperature and increasing time for the other 2 temperature conditions. Across all temperature conditions, there was a tendency toward better stability for NS Plus.

Results from the modeling comparing the 2 quantitative kits are shown in Table 2. Hemoglobin stability was similar, although there was a tendency toward longer stability for NS Plus at most temperatures. At ambient temperatures for both devices, the predicted stability, using time to less than 80% Hb recovery as the standard, was less than a week.

#### Study 2: Controlled Freeze-Thaw Study

As illustrated in Table 4, of the 3 kits tested, the NS Plus kit provided the most consistent and acceptable recovery after up to 3 freeze-thaw cycles. Recovery with the OC Diana kit declined steadily through the 3 freeze-thaw exposures. For the ulti med kit, the only qualitative test that could be included in this study, significant reduction in recovery was observed even after one cycle.

#### Study 3: Field Study

During transportation from the 3 distribution centers, inoculated devices were exposed to average temperatures of 18.4°C, 18.8°C, and 17.0°C. The lowest and highest temperatures encountered were −14°C and +26°C, respectively. The number of freeze-thaw cycles encountered during transport varied depending on the distribution center to which the devices were sent (one center had none, one center had 3, and one center had 1). Total transportation times varied from 5 to 12 days.

Table 5 summarizes the 2 analyses of agreement in the field study. Both NS Plus and OC Diana exhibited good agreement before and after transportation and with the expected result. Immunostics FIT and ulti med had good agreement before transportation, in keeping with their known analytic sensitivities, which should be able to detect the lower Hb concentrations in the inoculated devices. However, the agreement for both was poor between the before and after transportation tests, indicating degradation of Hb (ulti med more than Immunostics FIT). The Beckman ICT and Immunostics Guaiac had poor agreement before transportation, in keeping with their known poorer analytic sensitivities. However, agreement was good between the

### Table 2. Results From the Linear Regression Models Predicting the Stability of NS Plus and OC Diana Kits at Various Temperatures

| Temperature          | Period of Stability, d (95% CI) |
|----------------------|---------------------------------|
| Refrigerated         |                                |
| –2°C to 8°C          |                                 |
| (36°F–66°F)          |                                 |
| Elevated             |                                 |
| 20°C to 22°C         | 4.6 (3.3–5.7)                   |
| (68°F–72°F)          |                                 |
| NS Plusb             | 37.4 (32.3–46.5)                |
| OC Dianae            | 4.5 (1.6–6.9)                   |
| Frozen               |                                 |
| –20°C to –15°C       |                                 |
| (−4°F to 5°F)        |                                 |
| Refrigerated         |                                 |
| 2°C to 8°C           | >60d                            |
| (36°F–66°F)          |                                 |
| Elevated             |                                 |
| 45°C (113°F)         | 0°                              |

* Stability is defined using the number of days until the percentage of hemoglobin recovery was predicted to drop below 80% of the concentration at day 0.

** NS Plus, Alpharma Pharma Corporation (Chuo-ku, Osaka, Japan).

† OC Diana, Eiken Chemical Co., Ltd. (Taito-ku, Tokyo, Japan).

‡ Throughout the 60-day measurement period, the mean predicted hemoglobin recovery was more than 80%.

§ Throughout the 60-day measurement period, the mean predicted hemoglobin recovery was less than 80%.
before and after transportation tests, suggesting that Hb remained stable in the collection devices of those kits during transportation.

**DISCUSSION**

Using a standardized and rigorous methodology, we found important variations in Hb stability across 5 commercially available FIT kits suitable for use in organized CRC screening programs and 1 gFOBT kit currently in use in the Ontario, Canada, program. The Immunostics FIT appeared to be the best-performing qualitative kit, whereas, among the quantitative kits, NS Plus tended to perform slightly better than OC Diana. As temperatures increased above freezing, Hb recoveries decreased over time for all kits. Freezing samples resulted in a small initial loss of recovery in the collection devices of the quantitative kits. However, Hb concentration was then maintained across all time points. Some loss of recovery during the transport process was observed in the field study. The current study did not find uniformly better stability in FIT kits using dry collection devices (Immunostics FIT and Beckman ICT) compared with wet collection devices.

Although the lack of Hb stability over time, particularly at higher temperatures, is well documented, and seasonal variation in positivity rates has been previously reported, there are only a few reports comparing the Hb stability of different FIT kits. Interestingly, one of those reports also compared NS Plus and an OC product and, similar to our results, found that the NS Plus tended to exhibit slightly better stability. To our knowledge, Immunostics FIT has not been compared with other FITs for stability. In the current study, we report a decrease in Hb concentration of about 1.4%/d and about 2.1%/d at ambient temperatures (20°C–22°C) for the NS Plus and the OC Diana, respectively. One other study has also reported Hb instability for FITs at ambient temperatures, ranging from 1.7%/d to 7.8%/d, depending on the kit. Other studies have shown important differences in other performance charac-

---

*The median percentage of hemoglobin recovery compared with baseline (time 0 of NS Plus [a] and OC Diana [b]) over time at different temperatures. NS-Plus is manufactured by Alfresa Pharma Corporation (Chuo-ku, Osaka, Japan), and OC-Sensor Diana is manufactured by Eiken Chemical Co, Ltd (Taito-ku, Tokyo, Japan).*

---

Arch Pathol Lab Med—Vol 142, January 2018

Hemoglobin Stability in FIT and gFOBT—Catomeris et al 79
teristics for CRC screening across FIT kits. In sum, these studies highlight a key message: not all FIT kits are the same; as such, selection of FITs for use in organized screenings should include carefully consideration of data on performance and quality control.

There are few data on the effect of freezing on Hb stability in FIT collection devices, although a study from the Netherlands reported reduced FIT detection rates in the winter. We noted a drop in Hb concentration in the quantitative collection devices with freezing, but, once frozen, the recovery for the quantitative kits was relatively constant for up to 60 days. However, the effect of sequential freezing and thawing varied across the 3 kits we evaluated, with the NS Plus performing best. Our findings suggest that the process of freezing and thawing (which may occur during kit transport to the laboratory), and not necessarily the time frozen, may decrease Hb recovery.

Our field study was conducted outside a controlled laboratory setting, and as a result, there was considerable variation in important parameters, such as temperature, transportation time, freezing, and thawing. Although there may be limitations to this "real-world" approach in

| Temperature | Period of Stability, d (95% CI)¹b |
|-------------|----------------------------------|
| Frozen      |                                  |
| 80          | 0                                |
| 70          | >60 (NA)²                       |
| 50          | >60 (NA)²                       |
| MSS, d      | None stated                     |
| Refrigerated|                                  |
| 80          | >60 (NA)²                       |
| 70          | >60 (NA)²                       |
| 50          | >60 (NA)²                       |
| MSS, d      | None stated                     |
| Ambient     |                                  |
| 80          | 1.4 (NA–71.6)                   |
| 70          | 12.1 (NA–37.5)                  |
| 50          | 3.2 (4.6–NA)                    |
| MSS, d      | 7 (NA–8.5)                      |
| Elevated    |                                  |
| 80          | 2.4 (NA–8.5)                    |
| 70          | 5.1 (NA–10.7)                   |
| 50          | 9.3 (1.7–14.5)                  |
| MSS, d      | None stated                     |

Abbreviations: MSS, manufacturer’s stated stability at each temperature (none of the manufacturers described the methods used to measure stability); NA, not available.

¹ Results from the regression models under each temperature condition for all kits in which stability was defined by the number of days (95% CI) until the probability of remaining positive was predicted to drop below thresholds of 80%, 70%, and 50%.

² Hemoglobin NS-Plus, Alfresa Pharma Corporation (Chuo-ku, Osaka, Japan); OC-Sensor Diana, Eiken Chemical Co, Ltd (Taito-ku, Tokyo, Japan); Hema-Screen SPECIFIC, Immunostics, Inc (Eatontown, New Jersey); FOB Advanced, ultimed (Ahrensburg, Germany); Hemoccult ICT; Beckman Coulter, Inc (Brea, California); Hema-Screen Guaiac, Immunostics, Inc (Eatontown, New Jersey).

Table 4. Effect of Freeze-Thaw on Recovery of Hemoglobin in Various Fecal Immunochemical Test Collection Devices

| Freeze-Thaw Cycle | Quantitative Kits | Qualitative Kits |
|-------------------|-------------------|------------------|
|                   | NS Plusa          | OC Diana²        | FOB Advancedc |
|                   | Median Recovery, % (95% CI) |        | Samples Remaining Positive, No. (95% CI) |
| 0                 | 100               | 100              | 6               |
| 1                 | 91.7 (68.5–99.4)  | 74.9 (53.6–89.5) | 4 (67)          |
| 2                 | 95.1 (67.8–107.3) | 66.8 (53.0–79.7) | 1 (17)          |
| 3                 | 95.4 (64.4–106.7) | 57.6 (42.8–71.2) | 1 (0)           |

a Hemoglobin NS-Plus, Alfresa Pharma Corp (Chuo-ku, Osaka, Japan).
b OC-Sensor Diana, Eiken Chemical Co, Ltd (Taito-ku, Tokyo, Japan).
c FOB Advanced, ultimed (Ahrensburg, Germany).
generalizability, important lessons about Hb stability have been learned from other similar studies.\textsuperscript{13,25} Although there was some loss of recovery during the transport process in this field study, in spite of long transit times of up to 12 days and variations in temperature conditions (\textdegree{}C to \textdegree{}C), the 2 quantitative kits performed better (Hb remained stable in their collection devices across the full range of Hb concentrations tested) than the other kits. A number of studies have examined the potential clinical significance of exposure to higher temperatures\textsuperscript{13} and delayed sample returns.\textsuperscript{15,16} Unfortunately, we were not able to perform the field study during the summer when temperatures are expected to be hotter. Nevertheless, data from the current study suggest that the collection devices of the quantitative kits and some devices of the qualitative kits appear reasonably robust and may be suitable for postal transport in climates similar to that in Canada, at least in winter.

In wet collection devices, such as those used by some FIT kits, the rate of Hb degradation has been reported to be greater than that in dry collection devices, perhaps because of bacterial degradation of Hb in moist environments.\textsuperscript{13,14} The current study shows Hb stability in the collection devices of the Immunostics FIT and Immunostics Guaiac kits, which use a dry collection device, was, in general, longer than that of the other FIT kits. By contrast, Beckman ICT, which also uses a dry collection device, underperformed relative to some of the other kits using wet collection devices. These findings suggest that claims of superior sample storage capability may not be generalizable across all dry collection devices.

There are strengths and weaknesses to the current study. Because few data comparing Hb stability in collection devices across FIT kits are available, we attempted to address that issue with a novel, transparent, and easily reproducible methodology. To compare across kits, we established a definition of stability: for quantitative kits, time until Hb concentration decreased to less than 80%, a value selected based on internal studies of between-run precision; and for qualitative kits, time until the probability of remaining positive dropped below 3 prespecified thresholds. This approach is similar to that used by another study from the United Kingdom,\textsuperscript{17} although that study used a larger drop in Hb (\textgreater{}50% from baseline) to define a loss of stability. A second study from France\textsuperscript{18} simply reported the percentage of change in Hb concentration by day because the authors felt they could not use regression techniques because of differences among kits in the distribution of the data. The clinical significance of different thresholds is not known and likely varies with the Hb concentration of the fecal sample and the analytic sensitivity of the particular FIT, but use of a threshold allows a common reference with which to compare kits. Unlike the 2 other studies mentioned, we used regression methods appropriate to the distribution of the data, allowing us to obtain robust summary measures of stability over time; in addition, the use of generalized estimating equations accounted for multiple measures on the same kit.

One of the challenges we faced in developing this methodology was that the kits had significantly different analytic sensitivities. We addressed that issue by selecting fecal Hb concentrations appropriate to the reported analytic sensitivity of each kit. It was observed that some of the unspiked samples gave unexpected positive results with the Immunostics and Ultimed kits, which may indicate false-positive results in some of those samples. A second limitation of this study was that the design of the study offered no opportunity to correlate decreases in Hb concentrations with changes in polyp or cancer detection rates; therefore, we cannot comment on the clinical impact of the observed declines in Hb stability from the different kits. Data from recent studies have shown that there can also be important differences in clinical outcomes across FIT kits, even when using the same positivity threshold.\textsuperscript{26}

The use of a standardized methodology to compare Hb stability is a strength of this study. Although the clinical impact of the differences we report are not known, it may be reasonable

### Table 5. Agreement Between Before-Transportation and After-Transportation Results and Agreement With Expected Results in the Field Study\textsuperscript{a}

| Sample Hb Concentration | \textit{Expected Result}, µg Hb/g | N | Before | After | Before | After | Before | After | Before | After | Before | After | Before | After | Before | After | Before | After | Before | After | Before | After | Before | After | Before | After |
|-------------------------|---------------------------------|---|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Samples Producing Negative Results for Unspiked Samples, No. | 30 | 2 | 3 | 2 | 1 | 3 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Samples Producing Positive Results for spiked Samples, No. | 60 | 6 | 6 | 6 | 6 | 6 | 2 | 4 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 125 | 3 | 3 | 3 | 3 | 3 | 1 | 3 | 3 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 500 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 5 | 6 | 6 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 2000 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| Agreement with expected result, % | >100 | 97 | 100 | 97 | 93 | 97 | 73 | 90 | 50 | 67 | 67 | 57 | 57 | 57 | 57 | 57 | 57 | 57 | 57 | 57 | 57 | 57 | 57 | 57 |

Abbreviation: Hb, hemoglobin.

\textsuperscript{a} Hemoglobin NS-Plus, Alfera Pharma Corporation (Chuo-ku, Osaka, Japan); OC-Sensor Diana, Eiken Chemical Co, Ltd (Taito-ku, Tokyo, Japan); Hema-Screen SPECIFIC, Immunostics, Inc (Eatontown, New Jersey); FOB Advanced, Ultimed (Ahrensburg, Germany); Hemoccult ICT, Beckman Coulter, Inc (Brea, California); Hema-Screen Guaiac, Immunostics, Inc (Eatontown, New Jersey).

---

Hemoglobin Stability in FIT and gFOBT—Catomeris et al

Arch Pathol Lab Med—Vol 142, January 2018
to assume that kits that performed relatively poorly in our study may be likely to also do so in the real world. Our approach confers a benefit when comparing products for Hb stability because the reported stabilities in manufacturers' inserts almost certainly use various methodologies. In so doing, our study addresses a recent call for a standardized methodology to measure Hb stability in FIT collection devices.

In summary, the current report considers considerable variation in the performance of FITs for Hb stability. We found that the Immunostics FIT was the best performer among the qualitative kits. Both quantitative kits performed well, although the NS Plus appeared to have a slight edge over the OC Diana in our study. It is not known whether this difference will be maintained now that OC Diana has been updated with a new buffer. Because not all FIT methods are the same, the current study provides important performance data on the FIT methods evaluated that will assist CRC screening programs and providers in FIT selection. Finally, we describe a novel analytic approach to the comparison of Hb stability across kits, which can be used to address the well-recognized need for standardization in this area.

References
1. Kronborg O, Fenger C, Olsen J, Jorgensen OD, Sondergaard O. Randomised study of screening for colorectal cancer with faecal-occult-blood test. Lancet. 1996;348(9040):1467–1471.
2. Scholefield JH, Moss SM, Mangham CM, Whynes DK, Hardcastle JD. Nottingham trial of faecal occult blood testing for colorectal cancer: a 20-year follow-up. Gut. 2012;61(7):1036–1040.
3. Shaukat A, Mongin SJ, Geisser MS, et al. Long-term mortality after screening for colorectal cancer. N Engl J Med. 2013;369(12):1106–1114.
4. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. Gut. 2013;62(10):1637–1649.
5. Allison JE, Fraser CG, Halloran SP, Young GP. Population screening for colorectal cancer means getting FIT: the past, present, and future of colorectal cancer screening. J Hepatol. 2014;60(2):117–130.
6. Young GP, St John DJ, Winawer SJ, Rozen P; WHO (World Health Organization); OMED (World Organization for Digestive Endoscopy). Choice of faecal occult blood tests for colorectal cancer screening: recommendations based on performance characteristics in population studies: a WHO (World Health Organization) and OMED (World Organization for Digestive Endoscopy) report. Am J Gastroenterol. 2002;97(10):2499–2507.
7. Federici A, Giorgi Rossi P, Borgia P, Bartolozzi F, Farci S, Gausticchi G. The immunochemical faecal occult blood test leads to higher compliance than the guaiac for colorectal cancer screening programmes: a cluster randomized controlled trial. J Med Screen. 2005;12(2):83–88.
8. Hoffman RM, Steel S, Yee EF, Massie L, Schrader RM, Murata GH. Colorectal cancer screening adherence is higher with fecal immunochemical tests than guaiac-based faecal occult blood tests: a randomized, controlled trial. Prev Med. 2010;50(5–6):297–299.
9. Hol L, Wilshuizen M, van Ballegouwen M, et al. Screening for colorectal cancer: random comparison of guaiac and immunochemical faecal occult blood testing at different cut-off levels. Br J Cancer. 2009;100(7):1103–1110.
10. van Rossum LG, van Rijn AF, Lauer WJ, et al. Random comparison of guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. Gastroenterology. 2008;135(1):82–90.
11. Young GP, Sinatra MA, St John DJ. Influence of delay in stool sampling on fecal occult blood test sensitivity. Clin Chem. 1996;42(7):1107–1108.
12. Bövén LE, Fraser CG. Effect of delay in sampling on haemoglobin determined by faecal immunochemical tests. Ann Clin Biochem. 2008;45(pt 6):604–605.
13. Grazzini G, Ventura L, Zappa M, et al. Influence of seasonal variations in haemoglobin on faecal occult blood test for colorectal cancer screening: observational study from the Florence district. Gut. 2010;59(11):1511–1515.
14. Rabeeneck L, Rumble RR, Thompson F, et al. Fecal immunochemical tests compared with guaiac fecal occult blood tests for population-based colorectal cancer screening. Can J Gastroenterol. 2012;26(3):131–147.
15. van Roon AH, Hol L, van Vuuren AJ, et al. Are fecal immunochemical test characteristics influenced by sample return time?: a population-based colorectal cancer screening trial. Am J Gastroenterol. 2012;107(7):1385–1392.
16. van Rossum LG, van Rijn AF, van Oijen MG, et al. False negative fecal occult blood tests due to delayed sample return in colorectal cancer screening. Int J Cancer. 2009;125(4):746–750.
17. Carroll MRR, Piggett C, Pearson S, et al. National Health Service Bowel Cancer Screening Programme Hub: Evaluation of quantitative fecal immunochemical tests for haemoglobin. Guildford, United Kingdom: Guildford Medical Device Evaluation Centre; 2013. Poster presented at: EuroLabFocus 2014; October 7–10, 2014; Liverpool, United Kingdom.
18. Guitteny L, Guillaume E, Levillain R, et al. Analytical comparison of three quantitative immunochemical fecal occult blood tests for colorectal cancer screening. Cancer Epidemiol Biomarkers Prev. 2011;20(7):1492–1501.
19. Allison JE, Fraser CG, Halloran SP, Young GP. Comparing fecal immunochemical tests: improved standardization is needed [comment]. Gastroenterology. 2012;142(3):422–424.
20. Timmough J, Lansdorp-Vogelaar I, Allison JE. Faecal immunochemical tests versus guaiac faecal occult blood tests: what clinicians and colorectal cancer screening programme organisers need to know. Cancer. 2015;64(8):1327–1337.
21. McCulloch LJ, Nelder JA. Generalized Linear Models. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC; 1989. Monographs on Statistics and Applied Probability, vol. 37.
22. Langel KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986;73(1):13–22.
23. Rozen P, Waked A, Vilkin A, Levi Z, Niv Y. Evaluation of a desk top instrument for the automated development and immunochromatographic quantification of fecal occult blood. Menatl J Med. 2015;64(8):1327–1337.
24. van Rossum LG, van Oijen MG. Different seasons with decreased performance of immunochemical fecal occult blood tests in colorectal cancer screening [comment]. Gut. 2011;60(9):1303–1304; author reply 1304.
25. Chiang TH, Chuang SL, Chen SL, et al. Difference in performance of fecal immunochemical tests with the same hemoglobin cut-off concentration in a nationwide colorectal cancer screening program. Gastroenterology. 2014;147(6):1317–1326.
26. Hundt S, Haug U, Brenner H. Comparative evaluation of immunochromatographic fecal occult blood tests for colorectal adenoma detection. Ann Intern Med. 2009;150(3):162–169.
27.lee JK, Liles EG, Bent S, Levin TR, Corley DA. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. Ann Intern Med. 2014;160(3):171–180.
28. Park MJ, Choi KS, Lee YK, Jun JK, Lee HY. A comparison of qualitative and quantitative fecal immunochemical tests in the Korean nationwide colorectal cancer screening program. Scand J Gastroenterol. 2012;47(4):461–466.