Assessing the Size of Polyp Phantoms in Tandem Colonoscopies

CARLOS A. RUBIO1, CHARLOTTE M. HÖÖG2, OLLE BROSTRÖM2, JÖRGEN GUSTAVSSON2, MATS KARLSSON2, PER MORITZ2, ROBERT STIG2, OLA WIKMAN2, LARS MATTSSON3 and DOMENICO PALLI4

1Department of Pathology, Karolinska Institute University Hospital, Stockholm;
2Department of Gastroenterology, South Hospital, Karolinska Institute, Stockholm;
3Department of Production Engineering, The Royal Institute of Technology, Stockholm, Sweden; 4Molecular and Nutritional Epidemiology Unit, ISPO – Cancer Research and Prevention Institute, Florence, Italy

Abstract. Background: The size of colorectal neoplastic polyps is important for their clinical management. Materials and Methods: The size of 12 polyp phantoms was assessed in tandem colonoscopies carried out by 7 endoscopists differing in years of clinical endoscopical experience. The endoscopists measured, with (n=5) or without (n=2) the aid of open forceps, the largest diameter of 12 polyp phantoms. Measurements in two independent trials were compared with the gold standard-size assessed at The Department of Production Engineering, The Royal Institute of Technology. Results: In tandem trials, 99.4% (167/168) of the measurements underscored the gold standard size. In the 1st trial, the size in all 84 measurements was underestimated by –40% (range –34% to –45%) and in the 2nd trial the size in 83 of the 84 measurements was underestimated by –34% (range –24% to –42%). Neither the age of the participant, nor the years of experience with clinical endoscopy improved the results obtained. The participants significantly underestimated larger devices (≥20 mm) whereas the smallest “polyps” were also underestimated, but with a lower degree of inaccuracy. The absolute difference between the golden standard size and the mean of all measurements performed on each polyp in 167 out of 168 measurements followed a regular downward trend. The volume of the devices was one of the confounding factors in size assessment. When compared to the gold standard size, the larger the “polyp” size, the higher the degree of underestimation. This may be crucial considering that the risk for colorectal adenomas to shelter an invasive growth is 46%, for adenomas measuring ≥2 cm, a limit accepted as a guideline worldwide for the management of patients with large colorectal polyps. Conclusion: Considering the clinical implications of the results obtained, the possibility of developing a method that would allow the assessment of the true size of polyps in clinical colonoscopy, is being explored.

In 1887, Esmarch (1) pointed out that rectal adenomas could have a malignant potential. Forty years later, Feyter (2) confirmed these claims. Today it is generally understood that invasive carcinomas usually evolve from colorectal adenomas (3-13).

Thirty-five years ago it was demonstrated that the risk for colorectal adenomas to evolve into an invasive carcinoma correlated with the size of the lesion (3). In that classical work comprising 2506 colorectal adenomatous polyps, Muto et al. (3) found that the risk for adenomas measuring under 1 cm (largest diameter) to harbour an invasive carcinoma was 1%, for those measuring between 1 and 2 cm in diameter, the cancer risk increased to 10% and for those over 2 cm, to 46%. Pathologists worldwide (4-9) subsequently adopted these size limits as predictors of cancer risk and years later, these size limits were adopted by endoscopists (10-24).

To estimate the size of colorectal polyps, endoscopists often use as a reference, the distance between the opened blades of biopsy forceps. This method was applied to a large series of colorectal polyps (comprising 15989 polyps by the National Polyp Study Work Group (13,19), the Italian Multicentre Study Group (21), the Arizona Cancer Center Study (22), the South Korean Study Group (15) and more recently by Lieberman et al. (24). Patients were assigned groups based on the size of the poly(s) found at colonoscopy.

Using an artificial model, Margulies et al. (6) asked 6 gastroenterologists, 6 gastroenterology fellows and 7 untrained medical residents to measure ball bearings (3 mm to 19 mm) that were randomly inserted into a latex colon model. The size was estimated while being viewed with a
video colonoscope, with and without the aid of an open biopsy forceps. It was found that the estimated size by the 19 participants were consistently lower (13% to 29%) than the actual size for all groups, with and without forceps. These results were confirmed by Fennerty et al. (23) who utilized an in vitro latex colon model with 13 steel balls (as polyps) of different sizes sewn into an artificial sigmoidoscopy-teaching model. Eight experienced endoscopists estimated the size of these 13 polyps in two separate sessions, two weeks apart and the results were compared to the “actual polyp size”, defined as the largest diameter of each polyp, measured by three independent observers with a conventional millimetre ruler. The average of the three measurements was taken as the “actual polyp size”. The result showed that the estimates of the size of the polyps by these experienced endoscopists were 35% lower than “the true polyp size” for all polyps and for all endoscopists in the two sessions (23).

In a recent study it was found that the measuring of polyp phantoms with a conventional millimetre ruler by 22 pathologists (25) resulted in disparate inter and intra-observer estimations, at variance with the gold standard size (GSS) assessed at The Royal Institute of Technology, Stockholm. That was the first work ever published in which the size of polyp phantoms, assessed with a conventional millimetre ruler, was controlled with measurements carried out at a technological University. It was found that using a conventional ruler (the tool of pathologists worldwide) unacceptably high intra-observer and inter-observer variations in assessing the size of “polyp”-phantoms occurred. In a more recent work (26), the size of the same 12 phantom polyps was assessed by Computed Tomography (CT) and compared to the GSS. Size-assessment by CT was more reliable than that obtained with a millimetre ruler since the disparate individual deviation-values found with the latter method were avoided. The volume and the shape of the devices influenced size assessment of phantom polyps by CT.

The aim of the present work was to record the size of artificially confectioned polyp phantoms in tandem colonoscopies calculated by 7 endoscopists. Their results were subsequently compared to the GSS previously assessed at The Royal Institute of Technology, Stockholm.

Materials and Methods

The polyp phantoms. Twelve artificial “polyps” were used. A core consisting of round or flat hard buttons of various sizes was wrapped with papier-mâché, painted with a colour and subsequently varnished (Figure 1). After drying they were placed in individual vials labelled #1 to #12.

Colonoscopic procedure. The test was carried out in one of the rooms of the Endoscopic Unit. In that room, one endoscopist at a time measured the 12 polyp phantoms. A polyp phantom to be measured, chosen at random, was introduced into a cylinder measuring 43 cm long and 6 cm in diameter provided with a lock at one of the free ends (Figure 2). An Olympus colonoscope (PC 160, Tokyo, Japan) was subsequently introduced into the cylinder carrying a polyp phantom (Figure 3).

Measurements of polyp phantoms done at the Endoscopic Unit, Department of Gastroenterology. The size given by the endoscopist was registered by one of us (CAR) in a chart carrying the number of the “polyps” from #1 to #12.

After completing that procedure for the 12 devices, a second examination was done (tandem colonoscopy), this time changing the order of the number of the polyp phantoms, also at random. The procedure was subsequently repeated with the waiting endoscopist isolated to avoid possible bias by overhearing the results obtained by a previous participant. Five endoscopists chose to measure the 12 devices with the aid of a forceps and the remaining two, without that aid. Participants were asked to give the size of the largest diameter of the polyp phantoms in whole millimetres.

Measurements of polyp phantoms done at The Royal Institute of Technology. The measurement of the 12 devices was done by low force contacting metrology, at a temperature of 20°C ± 1°C. Held between the finger tips, each artificial “polyp” was rotated in a gap of two parallel metal surfaces of a micrometer screw. The distance between the surfaces was reduced until the largest diameter of the polyp caused a slight friction when turned around in the gap. A measurement series was performed in random order among the 12 artificial devices. The micrometer screw (Mitutoyo Digimatic MDC-25MJT) has a certified uncertainty of 0.0016 mm. Only the “polyp” with the largest diameter was measured with a calliper, as its size exceeded the micrometer screw measurement range. The Luna calliper has 0.1 mm uncertainty. The procedure was repeated every second day and after 5 measurements, the average and standard deviations for each sample was calculated. It is worth noting that the contacting low friction technique applied in this study cannot be used for elastic or soft devices. Then non-contacting optical techniques would be preferred. The limitations of such a procedure are the high number of projections required to determine the maximum diameter.

Statistical analysis. To evaluate intra-individual reproducibility, the Concordance Correlation Coefficient (CCC) between the first and second measure was calculated for each endoscopist participating in the study. The Pearson’s correlation coefficient (r) was also applied, to investigate the existence of a possible linear association between CCC and age, and between CCC and years of experience as endoscopist.

Each endoscopical measurement was compared to the absolute value provided by The Royal Institute of Technology (considered the gold standard size) and a percentage value was calculated. The mean of the percentages obtained in the 1st and the 2nd trials for each different pair endoscopist/polyp phantom was also calculated. In order to assess possible difference among the performances of the participants in the study, the Fisher’s F-test was done to analyze variance.

Results

The size polyp phantoms assessed at The Royal Institute of Technology (gold standard size). The results are presented in Table I. The results show that the standard deviation for
measurements of the largest diameter in the 12 devices was ≤0.05 mm when using the micrometer screw and ≤0.3 mm for the calliper. The difference in size in the 5 trials was non-significant.

The size given by endoscopists in the 1st and 2nd colonoscopy trials. The seven endoscopists that carried out two non-consecutive measurements of 12 polyp phantoms provided a total of 168 measurements.

Age and years of clinical experience of colonoscopic examination: The mean age of the participants was 52 years (range: 38-62) and the mean years of experience in the field was 14 years (range: 2-30) (Table II).

Measurements. The results of all the measurements given in the 1st and the 2nd trials, are shown in Table III.

Colonoscopic trial # 1: Table IV shows that the “polyp” with the largest deviation (in mm) from the gold standard for all participants was “polyp” #3 (mean –9.23 mm), followed by “polyp” #4 (mean –7.97 mm) and the lowest deviation was recorded for “polyp” #5 (mean –3.41 mm) followed by “polyp” #1 (mean –3.66 mm). Table IV also shows that participant B had the highest mean deviation (−9.19 mm) whereas participant C had the lowest mean deviation (−4.69 mm). While participant B had 10 years of endoscopical experience with colonoscopy (Table II), participant C had only 5 years of endoscopical experience. Whereas participant A (−5.19 mm) had 30 years of of endoscopical experience (Table II), participant E (−6.27 mm) had only 2 years of colonoscopic training. Apparently, the years of colonoscopic experience were of no help in correctly assessing the size of polyps at endoscopic examination in trial # 1. Expectedly, the two participants who did not use forceps in calculating size (participants B and F (−9.19 mm and −7.02, with 10 and 23 years of experience, respectively) had the highest mean deviation (Table IV).

Colonoscopic trial # 2: Table V shows that the device with the largest deviation from the gold standard was “polyp” #12 (mean –8.70 mm), followed by “polyp” #3 (mean –8.23 mm). The lowest deviation was recorded in “polyp” #5 (mean –1.98 mm) followed by “polyp” #1 (mean –3.38 mm). Table V also shows that participant G (with 23 years of experience) recorded the highest mean deviation (−6.77 mm), whereas participants A and C (with 30 and 5 years of experience, respectively) recorded the lowest mean deviations (−4.02 mm and −4.02 mm, respectively). Even in trial # 2, the years of colonoscopic experience were of no help in the correct assessment of the size of polyps at endoscopic examination. For participants not using forceps in calculating size (participants B and F) the mean deviation from gold standard was high (−6.69 mm –6.36 mm, respectively in Table V) but also by participant G (−6.77 mm), who used the aperture of the forceps in calculating the size of the devices.

Comparing size obtained in trials #1 and # 2 to the gold standard size. Table VI shows that the largest calculated difference in size between the 1st and the 2nd trials was given by participant B, with 10 years of experience (−7.94 mm) and the lowest by participant C, with only 5 years of clinical colonoscopic experience (−4.36 mm).

Whereas participant E, with only 2 years of experience had a measurement error of −5.69 mm between the 1st and the
2nd trials, participant G, with 23 years experience had an error of –6.06 between the 1st and the 2nd trials. Apparently the years of colonoscopic experience were of no help in reducing errors in the readings of the size of the devices in tandem trials. Table VII shows the poor overall concordance correlation coefficient (CCC) between the 1st and the 2nd trials. The intra-individual reproducibility indicated that only one CC value was lower than 0.70.

The Pearson’s r coefficients for the linear association between CC and age or years of experience as endoscopist resulted in non-significantly differences from the null value (not shown). The analysis of variance showed a significantly greater inter-individual reproducibility than the intra-individual one, with a F-test value of 11.54 (\(p<0.001\)).

**Discussion**

The results of this investigation demonstrated that 99.4% of 167 of a total of 168 measurements done by the 7 endoscopists in two tandem trials underscored the gold standard values. In the 1st trial, all 84 measurements underestimated the gold standard size by –40% (range –34% to –45%). In the 2nd trial, 83 of the 84 measurements underestimated the gold standard size by –34% (range –24% to –42%). Thus, the absolute difference between the golden standard size and the mean of all measurements performed on each polyp in 167 of 168 measurements followed a regular downward trend. The cause(s) responsible for the overall underestimation of the size of polyp phantoms by virtual colonoscopy in tandem trials remain elusive. Neither the age of the observer, nor the years of experience with clinical endoscopy improved the results obtained.

Some of the confounding factors that could help to explain the failure of endoscopists to adequately measure colorectal polyps may be: a) that endoscopists do not re-check their own initial polyp size with a second measurement, b) that endoscopists do not double-check the size obtained with another endoscopist, before the polyp is excised, c) that the distance between opened biopsy forceps used in colonoscopies may differ, when different forceps are used (for some 6 mm apart (23) and for others 8 mm apart (17), and d) that the blades of the forceps failed to open-wide, thus interfering with the calculation of the size of the polyp while extrapolating that aperture. Another, not previously envisaged confounding factor that might explain the failure of measuring polyps at endoscopy, is the large bulk of the polyp. In fact, “polyps” #3 and #12 were the largest (see Figure 1) and were the ones that were underscored with the largest deviation from gold standard by all participants. Participants significantly underestimated the larger devices, whereas the smallest “polyps” were also underestimated but with a lower degree of inaccuracy. Thus, the results of this work seem to suggest that when measuring polyps through a colonoscope, the larger the “polyp” size, the higher the degree of underestimation when compared to the gold standard size. This may be crucial considering that the risk for colorectal adenomas (histologically proven) measuring ≥2 cm, to shelter an invasive growth is 46% (3). That limit has been accepted worldwide as a guideline for the management of patients with “large” colorectal polyps.
Table III. Results of the two measurement (in mm) of 12 polyp phantoms done by the seven endoscopists participating to the study.

<table>
<thead>
<tr>
<th>No. polyp</th>
<th>1st A*</th>
<th>2nd A*</th>
<th>1st B</th>
<th>2nd B</th>
<th>1st C*</th>
<th>2nd C*</th>
<th>1st D*</th>
<th>2nd D*</th>
<th>1st E*</th>
<th>2nd E*</th>
<th>1st F</th>
<th>2nd F</th>
<th>1st G*</th>
<th>2nd G*</th>
<th>Size (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>8.52</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td>9</td>
<td>10</td>
<td>8</td>
<td>11</td>
<td>7</td>
<td>10</td>
<td>8</td>
<td>7</td>
<td>10</td>
<td>6</td>
<td>13.39</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>16</td>
<td>15</td>
<td>17</td>
<td>15</td>
<td>16</td>
<td>15</td>
<td>17</td>
<td>10</td>
<td>15</td>
<td>13</td>
<td>15</td>
<td>20</td>
<td>20</td>
<td>24.8</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>14</td>
<td>10</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>15</td>
<td>11</td>
<td>13</td>
<td>9</td>
<td>15</td>
<td>9</td>
<td>12</td>
<td>18.68</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>7</td>
<td>12</td>
<td>12</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>8.41</td>
</tr>
<tr>
<td>6</td>
<td>14</td>
<td>16</td>
<td>6</td>
<td>12</td>
<td>12</td>
<td>14</td>
<td>13</td>
<td>14</td>
<td>13</td>
<td>14</td>
<td>7</td>
<td>12</td>
<td>15</td>
<td>9</td>
<td>18.86</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>15</td>
<td>5</td>
<td>7</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>13</td>
<td>12</td>
<td>10</td>
<td>16.8</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>14</td>
<td>5</td>
<td>4</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>15</td>
<td>7</td>
<td>11</td>
<td>8</td>
<td>16.33</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>10.99</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>8</td>
<td>3</td>
<td>4</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>10.2</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
<td>14</td>
<td>7</td>
<td>15</td>
<td>10</td>
<td>12</td>
<td>9</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>6</td>
<td>14</td>
<td>10</td>
<td>16.58</td>
</tr>
<tr>
<td>12</td>
<td>20</td>
<td>20</td>
<td>17</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>18</td>
<td>18</td>
<td>20</td>
<td>20</td>
<td>17</td>
<td>20</td>
<td>15</td>
<td>15</td>
<td>27.7</td>
</tr>
</tbody>
</table>

*Used the biopsy forceps to calculate the size of polyp phantoms. Measurements of polyp size was performed at The Royal Institute of Technology.

Table IV. Deviation (compared to golden standard) of measurements (in mm) of 12 polyp phantoms given by 7 endoscopists participating in the study – 1st trial.

<table>
<thead>
<tr>
<th>Participant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>Endoscopist mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-0.52</td>
<td>-3.39</td>
<td>-9.80</td>
<td>-6.68</td>
<td>-1.41</td>
<td>-6.86</td>
<td>-4.80</td>
<td>-4.33</td>
<td>-1.99</td>
<td>-2.20</td>
<td>-6.58</td>
<td>-7.70</td>
<td>-4.69</td>
</tr>
</tbody>
</table>


Table V. Deviation (compared to golden standard) of measurements (in mm) of 12 polyp phantoms given by 7 endoscopists participating in the study – 2nd trial.

<table>
<thead>
<tr>
<th>Participant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>Endoscopist mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-1.52</td>
<td>-3.39</td>
<td>-8.80</td>
<td>-6.68</td>
<td>3.59</td>
<td>-4.86</td>
<td>-4.80</td>
<td>-4.33</td>
<td>-2.99</td>
<td>-2.20</td>
<td>-4.58</td>
<td>-7.70</td>
<td>-4.02</td>
</tr>
</tbody>
</table>

In daily praxis, endoscopists assess the size of colorectal polyps before removal, usually by comparing the aperture of the blades of the forceps and extrapolating that size to larger polyps, a method that was used here by 5 of the 7 participants. In the 1st trial, the two endoscopists that disregarded the aperture of the forceps in calculating the size of the “polyps”, recorded the highest underestimation size but in the 2nd trial the highest underestimation size was given by a endoscopist G with 23 years of clinical endoscopic experience, who used the aperture of the forceps to measure the devices. Thus, in the 2nd trial, the underscoring of the size of the “polyps” was not improved by the aid of the aperture of the forceps. The results obtained seem to substantiate the recent statement of Rex and Goldblum, namely that “colonoscopy is not perfect” (27).

In sum, the method of calculating “polyps” size by endoscopic examination as used in the present work, proved to be unreliable. And yet, the size of colorectal polyps given at “first glance” by the clinical endoscopist on duty is usually transferred to hospital records, results that subsequently are used in clinical publication concerning polyps at risk. Moreover, endoscopists together with pathologists, radiologists and gastroenterologists usually discuss at clinical conferences the strategy to be applied in the management of colonic polyps measuring 2 cm or more in diameter.

It should be understood that in clinical praxis, the work of endoscopists is very complex; they must focus not only on the needs of the patient during the examination, but also in the technical requirements, in the characteristics both of the colorectal mucosa as well as of the newly detected polyp (surface structure, colour, size and photographic documentation) before the polyp is removed with a forceps for subsequent histological evaluation at the Department of Pathology. Thus, in clinical praxis, size estimation is only one of several steps in the endoscopical examination of patients with newly detected colorectal polyps. This circumstance contrasts with that of the present work, in which the participants were only focused on the measuring of the devices through the colonoscope.

A method to improve the assessment of the size of endoscopically removed polyps is being developed (28). Alternative methods that would permit a better assessment of the size of clinical polyps before excision, that is at colonoscopical examination, should be explored.

Acknowledgements

Thanks are due to Dr. Saverio Caini, ISPO, Florence, for technical assistance.

References


Received October 8, 2008
Revised January 7, 2009
Accepted February 12, 2009