

Diagnostic accuracy and interobserver agreement of CT colonography (virtual colonoscopy)

P Pescatore, T Glücker, J Delarive, R Meuli, D Pantofflickova, B Duvoisin, P Schnyder, A L Blum, G Dorta

Abstract

Background and aims—Computed tomographic (CT) colonography or virtual colonoscopy (VC) is a non-invasive imaging method proposed for screening patients with colorectal neoplasias. Our aims were to study the diagnostic accuracy and interobserver agreement of VC for correct patient identification compared with conventional colonoscopy (CC).

Methods—This was a prospective study of 50 patients successively undergoing VC and CC. Multiplanar two dimensional CT images and three dimensional VC were constructed using surface rendering software and interpreted by two independent investigator teams. VC findings were compared with those of CC. Interobserver agreement was determined using kappa statistics.

Results—CC found 65 polyps in 24 patients. For identification of patients with polyps ≥ 10 mm, the sensitivity of VC was 38% and 63%, and specificity was 74% and 74% for teams 1 and team 2. Interobserver agreement was good (kappa 0.72). For patients with polyps of any size, the sensitivity of VC was 75% and 71%, and specificity was 62% and 69% for teams 1 and 2. Interobserver agreement was fair (kappa 0.56). Accuracy improved when comparing the results of the first 24 with the last 26 patients.

Conclusions—In our experience, VC had a low diagnostic value for identification of patients with colorectal neoplasias. Interobserver agreement for VC interpretation was fair. These results may be explained by software imperfections and a learning curve effect.

(Gut 2000;47:126-130)

Keywords: computed tomographic; colonography; colonoscopy; diagnostic accuracy; interobserver agreement

Screening for colorectal cancer is now widely recommended but raises important economic problems as the prevalence of precursor adenomas is about 25% after the age of 50 years.¹ The goal of screening should be detection of patients with these precursor lesions. Unfortunately, current screening methods are either poorly specific and sensitive (faecal occult blood testing, sigmoidoscopy) or too expensive and invasive (total colonoscopy).¹⁻³ Computed tomographic (CT)

three dimensional colonography or virtual colonoscopy (VC) was introduced in 1994 as a non-invasive rapid imaging method of the colon and rectum.⁴ On the basis of initial studies, VC appears to be an excellent diagnostic procedure compared with conventional colonoscopy for detection of polyps and carcinomas.⁵⁻⁸ However, an adequate and prospective evaluation of VC as a screening procedure has not been performed. Therefore, the aims of our study were to test the sensitivity and specificity of VC as well as interobserver variability, using commercially available software in a prospective series of 50 consecutive patients referred for diagnostic colonoscopy without recent prior morphological study of the colorectum.

Methods

This prospective study was conducted from March 1997 to March 1998 and was preceded by a series of VC test examinations to adjust the technical parameters. The study included 50 consecutive patients (31 men and 19 women) aged 50-85 years (mean 68 (8) years). The study design was approved by the local ethics committee and all patients gave informed written consent. All patients were aged >50 years and were referred for conventional colonoscopy (CC). Exclusion criteria were inflammatory bowel disease and refusal of consent. Indications for colonoscopy included: abdominal pain (n=11), iron deficiency anaemia of unknown origin (n=10), surveillance because of a personal history of colon polyps (n=10), haematochezia or positive occult faecal blood test (n=7), tumour search (n=7), or personal history of colorectal cancer (n=5). A partial colectomy had been performed in all patients with colorectal cancers.

After standard oral colonoscopy preparation (3 litres of polyethylene glycol; Fordtran, Streuli, Switzerland) patients underwent helical CT scanning (Advantage General Electrics, Switzerland). All patients received an intravenous injection of a musculotropic spasmolytic (N-butyl-hyoscine 20 mg; Buscopan, Boehringer Ingelheim, Germany). Patients were placed in the supine position and colon insufflation was performed with room air to maximal tolerance. A standard CT scout view was used to assess the degree of colonic distension. CT images were obtained during one or two breath-holds of 25-40 seconds, using 5 mm collimation with a table speed of 7.5 mm/s

Abbreviations used in this paper: CT, computed tomography; VC, virtual colonoscopy; CC, conventional colonoscopy.

Division of
Gastroenterology,
Centre Hospitalier
Universitaire Vaudois,
CH-1011 Lausanne,
Switzerland

P Pescatore
J Delarive
D Pantofflickova
A L Blum
G Dorta

Department of
Radiology, Centre
Hospitalier
Universitaire Vaudois,
CH-1011 Lausanne,
Switzerland

T Glücker
R Meuli
B Duvoisin
P Schnyder

Correspondence to:
Dr T Glücker, Service de
Radiodiagnostic et de
Radiologie Interventionnelle,
Centre Hospitalier
Universitaire Vaudois,
CH-1011 Lausanne,
Switzerland

Accepted for publication
11 January 2000

(pitch of 1.5). Axial CT images were reconstructed at 2.5 mm intervals. Scanning parameters were 200 mA and 110 kVp. We used a 512×512 matrix for image reconstruction. Scanning was in the craniocaudal direction. The two dimensional and three dimensional CT colonoscopy calculations were performed by down loading the data to a Sun Spark 20 computer workstation (Microsystems Mountain View, California, USA) equipped with commercially available software (Advantage Windows, Navigator, version 24/7/97, General Electrics, Switzerland). Using surface rendering techniques, multiplanar two dimensional views and a virtual fly-through (in retrograde and antegrade directions) of the colon were constructed and stored on optical CD for further viewing. The threshold value chosen for CT attenuation was -800 HU.

Conventional colonoscopy (CC) with either polypectomy of all resectable polyps or biopsy of all non-removable polyps and masses was performed immediately after helical CT by an independent experienced operator, unaware of the results of CT. Polyp size was determined by open biopsy forceps and verified immediately after polypectomy. VC interpretation was performed by two independent investigator teams each consisting of one radiologist and one gastroenterologist of similar seniority. In case of an incomplete CC, interpretation of VC concerned only the segments explored by CC. Both teams were blinded to the results of CC and the interpretation of the other team. However, intermediate data analysis with unveiling of the results of CC was performed after completion of VC interpretation for the first 24 patients; all participants had access to these results. The study protocol was not modified subsequently. A second analysis was performed after completion of VC interpretation for the remaining 26 patients. Thus the learning process was not studied on a case by case basis.

To compare VC with the reference method CC, the following prospective definitions were used: determination of sensitivity and specificity of VC for correct patient classification (with or without polyps). Polyps were grouped according to size: <10 mm or ≥10 mm in

diameter. VC findings were considered to correlate with CC findings when polyp size was identical ±3 mm, when polyp morphology was similar (sessile or pedunculated), and when VC located the polyp in the same colon segment as CC (rectum, sigmoid, descending colon, transverse colon, ascending colon, or caecum, respectively). A subset analysis according to polyp location in the six colon segments was performed. Statistical comparison of percentages was performed by Fisher's exact test. We evaluated interobserver agreement using the kappa coefficients (±standard error).⁹ Agreement is considered fair to good if kappa values are 0.4–0.75 and excellent if greater than 0.75. A kappa value of zero indicate absence of agreement; negative kappa values indicate disagreement.⁹

Analysis for individual polyp description has been reported elsewhere.¹⁰

Results

CC detected 65 polyps in 24 patients; 46 polyps were 1–5 mm in diameter, eight polyps measured 6–9 mm in diameter, and 11 polyps ≥10 mm in diameter. The distribution of polyps according to histology was: 35 adenomas and 11 hyperplastic polyps ≤5 mm; eight adenomas of 6–9 mm in diameter; seven adenomas ≥10 mm in diameter; and four carcinomas ≥10 mm (two of which were stenosing). Two colonoscopies were incomplete due to stenosing masses.

For identification of any particular *patient with polyps of any size*, VC had the following diagnostic values: sensitivity of 75% (95% confidence interval (CI) ±18%) and 71% (±18%) for teams 1 and 2 respectively; specificity of 62% and 69% (95% CI ±19%), respectively. Positive predictive values were 72% for both teams; negative predictive values were 64% and 68% for teams 1 and 2, respectively. When the analysis was restricted to patients carrying at least one or more *polyps ≥10 mm* in diameter, sensitivity was 37% (95% CI ±33%) and 62% (95% CI ±33%) for teams 1 and 2, respectively, and specificity was 74% (95% CI ±13%) for both teams. Positive predictive values were 21% and 31% for both teams; negative predictive

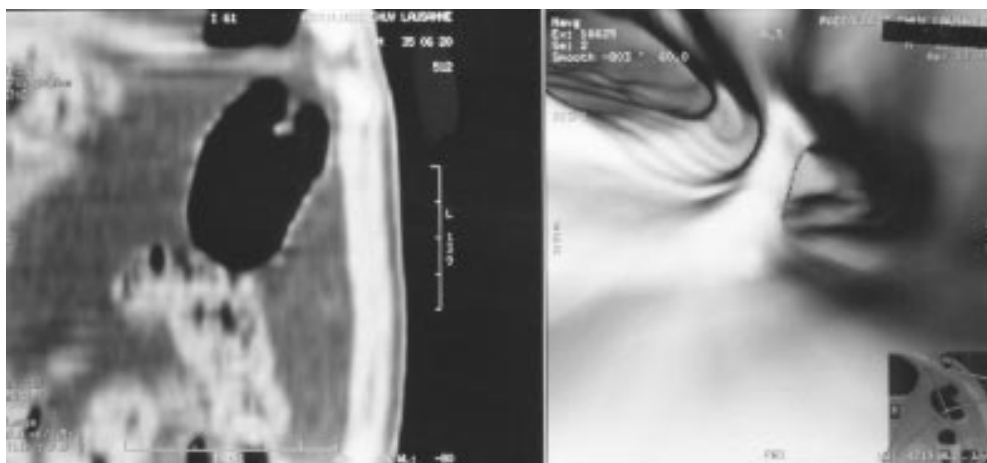


Figure 1 A false negative finding. A pedunculated polyp of 8 mm in diameter, seen on two dimensional reconstructions after review of the data and misinterpreted on the three dimensional view as a haustral fold.

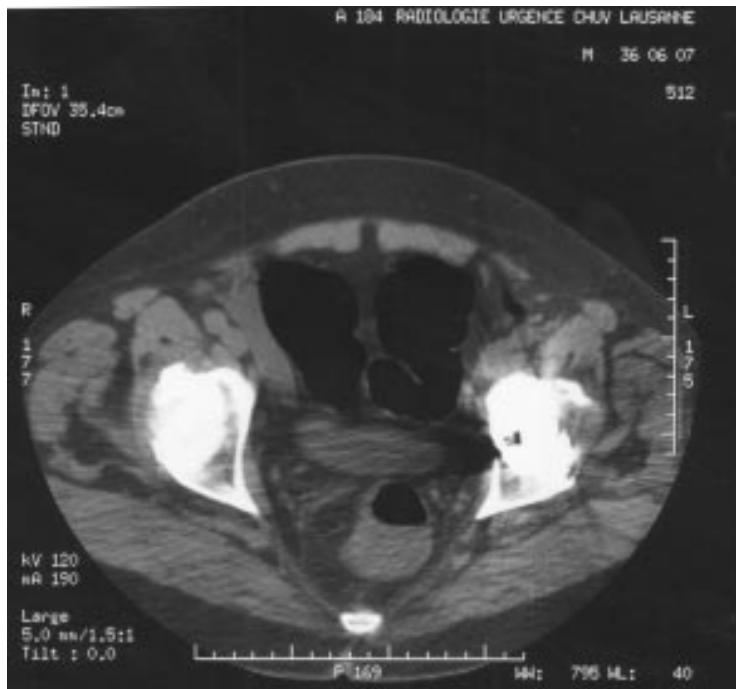


Figure 2 A false negative finding. Rectal carcinoma not correctly perceived by both teams on two dimensional reconstruction and hidden by fluid on three dimensional view.

values were 86% and 91% for teams 1 and 2, respectively. For patients with polyps <10 mm, values were: sensitivity 71% (95% CI \pm 19%) for both teams and specificity 59% (95% CI \pm 18%) and 69% (95% CI \pm 17%) for teams 1 and 2, respectively. Positive predictive values were 55% and 62% for both teams; negative predictive values were 74% and 77%, respectively.

False negative findings for patients with polyps \geq 10 mm occurred in six and three cases, respectively, according to teams 1 and 2. In an attempt to explain the low sensitivity values, we analysed all false negative results for polyps \geq 10 mm for a total of six patients with 11 lesions. Team 1 had missed seven, including three of four carcinomas. The reasons for failure to detect these seven lesions were mostly perceptible errors (four cases), explained by inadequate analysis of the two dimensional CT images (three cases) and polyp masked by fluid (one case). The three remaining polyps could not be found on review of the whole data set and repeated multiplanar reconstructions. Team 2 had missed four, including one of four carcinomas. Examples of false negative findings are shown in figures 1 and 2.

Specificity and sensitivity values for identification of patients with polyps of any size were analysed separately according to the *study periods*. For patients 1–24, sensitivity was 100% and 92%, and specificity was 42% and 58% for teams 1 and 2, respectively. For patients 25–50, sensitivity was 50% and specificity was 79% for both teams. Differences in sensitivity according to the study periods were statistically significant for teams 1 ($p=0.01$) and 2 ($p=0.04$). Differences in specificity according to study periods were not statistically significant ($p=0.1$ and 0.4 for teams 1 and 2, respectively).

Sensitivity values for individual polyp detection were analysed separately for *anatomical*

location and for teams 1 and 2, and were: rectum 0% and 0%; left colon 32% and 32%; transverse colon 63% and 50%; and right colon 33% and 25%, respectively. Specificity for detection of a particular polyp could not be calculated.

Kappa values were 0.56 (0.12) for patients with polyps of any size and 0.72 (0.10) for patients with polyps \geq 10 mm in diameter.

Discussion

This study of the diagnostic accuracy of VC was performed in a series of 50 consecutive patients without proven colorectal pathology referred for diagnostic colonoscopy whereas previously published studies were done under experimental conditions (artificial polyps, animal models)^{5 11–13} or included patient groups selected on the basis of recent morphological examinations.^{6 14–16} Moreover, in some of the latter studies the investigators were not blinded to the results of the examinations.^{6 15 16} Several other groups reported their initial experience in abstract form.^{17–21} Recently, two studies in partially asymptomatic patients, more closely resembling a screening population, were reported.^{22 23} All of these studies, including our own, compared VC with CC. Although CC is not a true “gold standard” for polyp detection, as shown by high miss rates of adenomas after back to back colonoscopy,²⁴ it is the best method of reference.²⁵ Most studies reported excellent sensitivities of VC for detection of patients with lesions greater than 10 mm in size.^{17 19 20 22} Hara *et al* correctly identified patients carrying polyps of more than 10 mm in diameter with a sensitivity of 75% and a specificity of 90%.⁶ Fenlon *et al* reported a sensitivity of 96% and a specificity of 96% using the same criterion.²² Preliminary results by Pineau *et al* were comparably good.²⁰ In contrast, lesions <5 mm were commonly diagnosed with a low accuracy.^{6 14–23} This is because of insufficient spatial resolution of the reconstruction algorithms.¹⁵ A notable contrast to these enthusiastic results is the study by Rex and colleagues²³ who reported low sensitivities for polyp detection. In this study sensitivities for correct identification of patients with any polyp >20 mm, with polyps of 10–19 mm, and those of 6–9 mm were 75%, 83%, and 43%, respectively. These values are insufficient for screening purposes. A high rate of false positive findings has also been reported by some investigators^{18 19} which in a screening programme would eventually result in performance of unnecessary complementary CC.

Our study was unique in that it involved data interpretation by two independent observer teams with ratings for interobserver agreement, and assessment of a learning process. We assumed that the most important goal of VC was correct patient selection for performance of CC. The results of our study do not agree with most published series. For patients with polyps \geq 10 mm, our two observer teams found low sensitivities of 38% and 63%; specificity was only 74% for both teams. The low sensitivity might be explained by three factors: firstly,

preparation quality was often suboptimal.¹⁰ This becomes especially obvious after analysis of miss rates according to anatomical location. Thus all lesions in the rectum were missed, mostly as a result of fluid persistence in the rectum. Secondly, the resolution capacity of the software may have been insufficient. Indeed surface rendering software (as used in this study) appears to be less reliable than more complex, time consuming, expensive volume rendering software.²⁶ Problems in image reconstruction have been reported by another group²³ who were also unable to find lesions after a second analysis of the data set. In this series, however, mostly flat adenomas and not stenosing masses were missed. Finally, as our study reflected early experience with the method, there was a lack of expertise, in particular in the first part of the study.

The low specificity reported was due to a learning process reflected by the fact that the false positive findings decreased after analysis of the first 24 patients. Accordingly, VC specificity was very modest for patients 1–24 but considerably improved for patients 25–50. However, sensitivity decreased as specificity improved, not only because better experience led to underdiagnosis of the clinically less relevant small polyps but also due to a lack of diagnosis of large lesions. This finding also explains the paradoxical better sensitivity for detection of patients with small adenomas (<10 mm). The explanation lies in the high rate of false positive findings—that is, the overall low specificity of the method. We prospectively divided our data into two groups to assess a possible learning effect. Although this method does not reflect a true learning process on a case by case basis, it nevertheless proved useful to document changes in data perception with increasing experience.

It could be argued that underdiagnosis of small polyps is clinically irrelevant as only 1.3% of polyps <10 mm are malignant.^{27–28} However, at least 5% of polyps <10 mm contain high grade dysplasia.^{27–29} Moreover, the goal of screening should be identification not only of early stage carcinomas but of precursor adenomas of any size. The most important characteristic of a screening procedure should be optimal sensitivity, even if specificity is relatively low.^{1–30} In addition, small size carcinomas and flat adenomas, which probably have a high prevalence even in Western countries,^{31–32} are undetectable with current VC technology.²³

The fair interobserver agreement reported reflects the fact that interpretation of VC depends on the observer's impressions and experience. A solution to this problem could be the advent of automated interpretation programmes which recognise polypoid formations of a certain size or tissue characteristic.³³ No clinical results have been published with these devices and there remains lack of standardisation as well as numerous pitfalls for computers.³⁴

On the basis of this study, the following improvements in VC are proposed: optimal bowel preparation; observing the two dimensional image set first before switching to the

more time consuming study of three dimensional images^{11–15}; and scanning in both prone and supine positions to mobilise fluid deposits.³⁵ Volume rendering or perspective volume rendering may be better than surface rendering programs but these require a more complex computer workstation.²⁶

In conclusion, in our experience VC does not yet appear to be suitable for colorectal cancer or polyp screening.³⁶

Presented in part at the Annual Meeting of the American Gastroenterological Association, New Orleans, May 18–22, 1998 (*Gastroenterology* 1998;114:A662).

- 1 Winawer SJ, Fletcher RH, Miller L, et al. Colorectal cancer screening: clinical guidelines and rationale. *Gastroenterology* 1997;112:594–642.
- 2 Eddy DM. Screening for colorectal cancer. *Ann Intern Med* 1990;113:373–384.
- 3 Lieberman DA, Smith FW. Screening for colon malignancy with colonoscopy. *Am J Gastroenterol* 1991;86:946–51.
- 4 Vining DJ, Gelfand DW. Noninvasive colonoscopy using helical CT scanning, 3D reconstruction and virtual reality. SGR Scientific Program, Maui, Hawaii, Feb 1994:70 (abstract).
- 5 Hara AK, Johnson CD, Reed JE, et al. Detection of colorectal polyps by computed tomographic colography: feasibility of a novel technique. *Gastroenterology* 1996;110:284–90.
- 6 Hara AK, Johnson CD, Reed JE, et al. Detection of colorectal polyps with CT colography: initial assessment of sensitivity and specificity. *Radiology* 1997;205:59–65.
- 7 Johnson CD, Hara AK, Reed JE. Computed tomographic colography (virtual colonoscopy): a new method for detecting colorectal neoplasms. *Endoscopy* 1997;29:454–61.
- 8 Johnson CD, Ahlquist DA. Computed tomography colography (virtual colonoscopy): a new method for colorectal screening. *Gut* 1999;44:301–5.
- 9 Fleiss JL. The measurement of interrater agreement. In: *Statistical methods for rates and proportions*. New York: John Wiley, 1981:212–36.
- 10 Glücker T, Delarive J, Pescatore P, et al. Evaluation de l'endoscopie virtuelle pour la recherche de lésions coliques. *J Radiol* 1998;79:A110 (abstract).
- 11 Hara AK, Johnson CD, Reed JE, et al. Colorectal polyp detection with CT colography: two- versus three-dimensional techniques. *Radiology* 1996;200:49–54.
- 12 Dachman AH, Lieberman J, Onsis RB, et al. Small simulated polyps in pig colon: sensitivity of CT virtual colography. *Radiology* 1997;203:427–30.
- 13 Beaulieu CF, Napel S, Daniel BL, et al. Detection of colonic polyps in a phantom model: implications for virtual colonoscopy data acquisition. *J Comput Assist Tomogr* 1998;22:656–63.
- 14 Kay CL, Kulling D, Hawes RH, et al. CT-based virtual endoscopy—correlation with colonoscopy in the detection of space-occupying lesions of the colon. *Radiology* 1997;203:195 (abstract).
- 15 Royster AP, Fenlon HM, Clarke PD, et al. CT colonoscopy of colorectal neoplasms: two-dimensional and three-dimensional virtual-reality techniques with colonoscopic correlation. *AYR* 1997;169:1237–42.
- 16 Fenlon HM, Nunes DP, Clarke PD, et al. Colorectal neoplasm detection using virtual colonoscopy: a feasibility study. *Gut* 1998;43:806–11.
- 17 Akerkar GA, Hung RK, Yee J, et al. Sensitivity and specificity of virtual colonoscopy for detection of colorectal neoplasia. *Gastroenterology* 1999;116:A44 (abstract).
- 18 Bethge N, Breitkreuz C, Harms S, et al. Prospective, controlled, double-blinded trial of virtual colonoscopy in occult GI bleeding. *Gastroenterology* 1999;116:A378 (abstract).
- 19 Morrin M, Farrell RJ, Kruskal JB, et al. Prospective comparison of virtual CT colography in patients undergoing elective colonoscopy. *Gastroenterology* 1999;116:A461 (abstract).
- 20 Pineau BC, Mikulaninincek C, Vining DJ. Ability of virtual colonoscopy to detect patients with colorectal polyps. *Gastroenterology* 1999;116:A485 (abstract).
- 21 Wong BCY, Chan JKF, Lam SK, et al. Virtual colonoscopy with 3-dimensional CT scan reconstruction for detection of colonic lesions in Hong Kong Chinese. *Gastrointest Endosc* 1999;49:AB161 (abstract).
- 22 Fenlon HM, Nunes DP, Schroy PC, et al. A comparison of virtual and conventional colonoscopy for the detection of colorectal polyps. *N Engl J Med* 1999;341:1496–503.
- 23 Rex D, Vining D, Kopecky KK. An initial experience with screening for colon polyps using spiral CT with and without CT colography (virtual colonoscopy). *Gastrointest Endosc* 1999;50:309–13.
- 24 Rex DK, Cutler CS, Lemmel GT, et al. Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. *Gastroenterology* 1997;112:24–8.
- 25 Wayne JD. What is a gold standard for colon polyps? *Gastroenterology* 1997;112:292–4.
- 26 Rubin GD, Beaulieu CF, Argiro V, et al. Perspective volume rendering of CT and MR images: applications for endoscopic imaging. *Radiology* 1996;199:321–30.

- 27 Morson BC. The polyp cancer sequence in the large bowel. *Proc R Soc Med* 1974;74:451-7.
- 28 Muto T, Bussey HJR, Morson BC. The evolution of cancer of the colon and rectum. *Cancer* 1975;36:2251-70.
- 29 O'Brien MJ, Winawer SJ, Zauber AG, et al. The National Polyp Study Workgroup. The natural polyp study. Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas. *Gastroenterology* 1990;98:371-9.
- 30 Bond JH, Mandel JS, Church TR. Role of fecal occult-blood testing. *N Engl J Med* 1998;339:774-5 (letter).
- 31 Fuji T, Rembacken BJ, Dixon MF, et al. Flat adenomas in the United Kingdom: are treatable cancers being missed? *Endoscopy* 1998;30:437-43.
- 32 Trecca A, Fuji T, Kato S, et al. Small advanced colorectal adenocarcinomas: Report on three cases. *Endoscopy* 1998;30:493-5.
- 33 Vining D, Ge Y, Ahn D, et al. Enhanced virtual colonoscopy system employing automatic detection of colon polyps. *Gastroenterology* 1998;114:A698 (abstract).
- 34 Hara AK, Johnson CD, Reed JE. Colorectal lesions: evaluation with CT colography. *Radiographics* 1997;17:1157-67.
- 35 Chen SC, Lu DSK, Hecht JR, et al. CT colonography: value of scanning in both the supine and prone positions. *AJR* 1999;172:595-9.
- 36 Bond JH. Virtual colonoscopy—Promising, but not ready for widespread use. *N Engl J Med* 1999;341:1540-2.