Comparison of Rectal Mucosal Proliferation Measured by Proliferating Cell Nuclear Antigen (PCNA) Immunohistochemistry and Whole Crypt Dissection

Sharon C. Murray, Robert S. Sandler, Temitope O. Keku, Cynthia M. Lyles, Robert C. Millikan, Shrikant I. Bangdiwala, Lawrence L. Kupper, Wuhan Jiang, and Martin H. Ulshen

Departments of Medicine, Epidemiology and Biostatistics, and Center for Gastrointestinal Biology and Disease, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7080

Abstract
Rectal mucosal proliferation has been promoted as an intermediate marker for risk of colorectal neoplasia. Proliferating cell nuclear antigen (PCNA) immunohistochemistry has become a standard method to measure cell proliferation. Whole-crypt dissection may provide a technically simpler method for determining proliferation within an entire crypt. We conducted a study to assess the reliability (reproducibility) of whole-crypt dissection in 10 subjects, and we compared whole-crypt dissection to PCNA immunohistochemistry in 91 subjects. Reliability of whole-crypt dissection with the subject as the unit of observation was excellent. The intraclass correlation coefficient for subjects was 0.93. Biopsy-to-biopsy reliability was lower ($r = 0.86$) and crypt-to-crypt reliability lower still ($r = 0.35$). There was poor correlation between measures of proliferation index using the two techniques (Kendall’s tau = 0.13; $P = 0.08$). Compartment analysis based on the percentage of crypt-to-crypt reliability lower still ($r = 0.35$).

Introduction
Cellular proliferation in the normal large intestine is generally confined to the base of colonic crypts. Absorptive and goblet cells migrate upward from putative stem cells at the crypt base (1). An expansion of the proliferative zone has been demonstrated in humans with diseases that predispose to cancer (such as familial polyposis coli) and in rodents treated with chemical carcinogens (2). Expansion of the proliferative zone has also been reported in the mucosa of patients with sporadic adenomas, nonfamilial colon cancer, and hereditary nonpolyposis colorectal cancer (3-5). Proliferation measures might, therefore, be used as biomarkers to identify individuals at increased risk for neoplasia or could be used to monitor the effects of intervention measures to reduce risk (6). Indeed, proposed chemopreventive agents have been shown to decrease the proliferative index in both rodents and humans (7, 8, 9).

A number of methods have been used to measure cellular proliferation in the large bowel. The determination of proliferation index by tritiated thymidine autoradiography has been used for many years but has some disadvantages (7, 10). The technique is somewhat time consuming and requires radioactive material. Bromodeoxyuridine is an analogue of thymidine that is incorporated into the DNA of replicating cells during S-phase of the cell cycle (11, 12, 13). This technique is faster than autoradiography and has been shown to have good correlation with tritiated thymidine incorporation (4). PCNA is a cell cycle-associated protein that is maximally synthesized during S-phase (14). Identifying PCNA is a technically simpler procedure because it does not require an incubation step. Because of ease of use and comparability to more established methods of measuring cell proliferation, PCNA immunohistochemistry (henceforth PCNA) may largely supplant other techniques in large-scale or multicenter studies.

All of these methods involve calculating the proportion of labeled cells from a cross-section of a colonic crypt. The whole-crypt dissection method measures the number and location of mitoses in an entire microdissected crypt. The number of mitoses in an entire crypt has been termed the WCMC. There are several theoretical advantages to this technique. Biopsies can be placed directly into fixative (Carnoy’s solution) and stored in alcohol until analysis; there is no need to maintain tissue viability or for incubation: tissue orientation is unimportant; and hundreds of crypts can be easily counted from a single endoscopic biopsy (15). The method is less time consuming than PCNA. Preliminary work in fasted and refed rats has suggested that the results of crypt dissection are comparable to those of the bromodeoxyuridine method (15). The method has not been extensively used or validated in average-risk human subjects.

If WCMC were highly reliable and comparable to PCNA LI, it might offer a much simpler biomarker for use in intervention studies designed to reduce the risk of colorectal malignancy. The 3-fold purpose of this study was: (a) to assess the reproducibility of the whole-crypt dissection method in human tissue obtained with standard endoscopy forceps; (b) to compare WCMC to PCNA LI; and, (c) to develop sample size nomograms that could be used to plan clinical studies.

Received 11/21/94; accepted 6/13/95.

The abbreviations used are: PCNA, proliferating cell nuclear antigen; WCMC, whole-crypt mitotic count; LI, labeling index.
Materials and Methods

Biopsies were obtained from consenting patients who were referred to the University of North Carolina Hospitals for a clinically indicated colonoscopy. Data are presented on 91 consecutive patients for whom both WCMC and PCNA LI were calculated: 37 white males, 37 white females, 4 black males, and 13 black females. Six biopsies were taken from the rectum at the start of the procedure using standard endoscopy forceps at approximately 8–12 cm from the anal verge. Each biopsy was transferred to bibulous paper (Fisher Scientific Co., Raleigh, NC) with special care not to stretch or tear each specimen. Four specimens were placed in Steinberg’s Modified Eagle’s Medium for PCNA, and two specimens were placed in Carnoy’s fixative for later crypt dissection. All specimens for PCNA were processed within 30 min.

The specimens for PCNA were fixed in 70% ethanol before processing. Paraffin blocks were prepared according to routine histological procedures, and care was taken not to exceed 60°C. Five sections were placed on glass slides precoated with poly-t-lysine (Sigma Chemical Co., St. Louis MO). The sections were 5-μm thick and were taken at least 50 μm apart so that each section would contain different crypts. The sections were incubated for 12–20 h at 4°C using a rat testis and human colon tissue as positive control slides. The sections were rehydrated through graded ethanol concentrations. Slides were incubated in Schiff’s reagent and incubated in 45% HCl at 60°C for 10 min in a water bath. The tissues for whole-crypt dissection were transferred from Carnoy’s fixative to 70% ethanol after 2.5–3.0 h. Prior to staining, the tissue was rehydrated in 5% and 25% ethanol, then hydrolyzed in 1 M HCl at 60°C for 10 min in a water bath. The biopsies were then placed in Schiff’s reagent and incubated in the dark for 45–60 min at room temperature. A small piece of biopsy was then cut and placed on a microscope slide with 45% acetic acid. The tissue was gently teased apart under a dissecting microscope, then covered with a coverslip. The coverslip was gently tapped until the crypts began to separate into discrete crypts. The dissected crypts were examined under a compound microscope at ×200. To be scored, the top and bottom of the crypt had to be seen. Broken crypts were not acceptable.

Mitotic cells were identified by nuclear clumping, focusing through the entire crypt. Each crypt was also traced on paper using a drawing attachment (Olympus Corporation, Lake Success, NY). All mitoses were drawn within the crypt. For straight crypts, the height of each crypt was measured from the lowest point to the center of a line drawn through the apex. Curved crypts were measured using a flexible curve. The position of each mitosis was measured from the crypt base. A total of 10 crypts were counted from each of the two biopsies. If there were not 10 countable crypts in the first piece taken from a given biopsy, another piece was taken.

All data were entered and then independently sight-verified. Range checks and checks for impossible values were also performed.

Reproducibility. The intrareader reliability of the crypt dissection method was assessed using data from ten patients whose slides were read twice by a masked observer. Scatter plots were generated comparing the values recorded for the first reading and second reading on a per crypt, per biopsy, and per patient basis. Intraclasse correlation coefficients were calculated in each of the three cases, using the natural log of the mitotic cell count, as a way of clarifying the degree of agreement between the two different readings of the same slides. The natural log scale was chosen because the distribution of the number of cells in mitosis tends to be skewed to the right.

Correlation between PCNA and Crypt Dissection Indices. Both PCNA and crypt dissection measurements were obtained for 91 patients. WCMC for each individual was calculated as an average of the number of cells undergoing mitosis in each of 20 crypts, 10 from each of 2 biopsies. PCNA LI was calculated for each individual as the average of the proportion of labeled cells from all scorable crypts in each of two biopsies. A scatter plot comparing the subject mean values of PCNA LI to WCMC was generated using SAS/GRAPH (SAS Institute, Cary, NC). Kendall’s τ, a nonparametric measure of correlation, and its associated P were calculated.

Compartment Analyses. Proliferation is not uniformly distributed throughout the length of colonic crypts. To evaluate the distribution of proliferation, we divided crypts into four quartiles. For both measures, we calculated the proportion of all labeled cells appearing in each quartile. For PCNA LI, this proportion was calculated for each crypt column as the number of labeled cells in the particular quartile divided by the total number of labeled cells in the crypt column. It was necessary to work with crypt columns rather than crypts because the number of cells in each column of the crypt was not always the same. The average proportion of labeled cells in each quartile was computed for each biopsy, and a score for each subject was computed as the average of the biopsy scores. The denominator used in computing this score differs from that usually seen in the literature for PCNA LI compartment analysis, i.e., the total number of cells in the compartment. Instead, the number of labeled cells in the entire crypt column was chosen as the denominator so that the PCNA compartment scores would be comparable to those obtained using crypt dissection, where we do not know the total number of cells in each compartment.

For crypt dissection, the height of the crypt was used to divide the crypt into quartiles. The number of labeled cells in each quartile (where quartile position is determined by the distance of the labeled cell from the base of the crypt) divided by the total number of labeled cells in the entire crypt was calculated for each quartile in each crypt. These quartile scores were then averaged for each of the two biopsies, and the mean of the biopsy scores was taken as the subject score. Correlation
indices were calculated comparing the average percentage of labeled cells appearing in each quartile for PCNA and crypt dissection. The overall mean of the subject scores was calculated for each of the quartiles for both PCNA and whole crypt dissection. These overall group mean scores were graphically compared using a histogram.

**Sample Size Calculations.** Minimum sample sizes necessary to attain 90% power for detecting a 25% difference in the proliferation measure between two groups were calculated for various sample size combinations of subjects, biopsies, and crypts for whole-crypt dissection. These calculations were based on a mixed-effects model using the log of the proliferation measure as the response.

The model used included a fixed effect due to group and random effects due to subject, biopsy, and crypt. Under this model, the log of the proliferation measure is assumed to be normally distributed. The hypothesis being tested that there is no difference between the two groups (H$_{0}$: $\delta = 0$) and the alternative hypothesis is two-sided (H$_{A}$: $\delta \neq 0$). The number of subjects, $n$, required per group is:

$$n = \frac{2(\sigma^2 + \sigma^2_s + \sigma^2_c)}{\delta^2},$$

where $\delta$ represents, in log-scale, the minimum absolute detectable difference between groups, $\alpha$ is the type I error rate, $B$ is the number of biopsies per subject, and $C$ is the number of crypts per biopsy. Also, $\sigma^2$, $\sigma^2_s$, and $\sigma^2_c$ are the components of variation due to subject differences, differences between biopsies within subjects, and differences between crypts within biopsies, respectively. The type I error rate was set to $\alpha = 0.05$, and a group difference of $\delta = \ln(0.75)$, which in log-scale corresponds to a 25% difference in the proliferation measure, was used.

Note that if one assumes that $\sigma^2_s$ and $\sigma^2_c$ are 0 in the sample-size formula given above, the result is the familiar formula for determining the required sample size when comparing two groups using a $t$ test. The simpler formula ignores the variation inherent in measurements taken on the same individual, leading to an underestimate of the required sample size.

**Results**

**Reproducibility.** For the crypt dissection data on the 10 subjects whose slides were read twice, the mean WCMC for the data obtained on the first reading was 4.85 (SD 2.83). On the second reading, the mean WCMC was 4.88 (SD 3.17). In Fig. 1, the values obtained upon a second reading of the data are plotted against the original scores on a per-crypt basis, a per-biopsy basis, and a per-subject basis. Note that the number of cells in mitosis in a crypt ranged from 1 to 24 (Fig. 1A). Per biopsy, the average number of cells in mitosis ranged from 2.1 to 12.9 (Fig. 1B), while the per-subject index ranged from 2.6 to 10.1 (Fig. 1C). It is clear from the graphs that the reproducibility of this index is better when looked at in terms of biopsy-to-biopsy values than in terms of crypt-to-crypt values and better still when the values are examined per subject. Since the biopsy score is an average of crypt scores and the subject score is an average of the biopsy scores, this result is as expected. Calculation of the intraclass correlation coefficient, which measures the degree to which the observations deviate from the 45° line of equivalency, gives 0.35 for the crypt-to-crypt comparison, 0.86 for the biopsy-to-biopsy comparison, and 0.93 for the subject-to-subject comparison.

**Correlation between PCNA and Crypt Dissection Indices.** For the 91 subjects for whom both PCNA and crypt dissection measures were available, the overall mean PCNA LI was 0.066.
Comparison of PCNA and Crypt Dissection

Fig. 2. Scatter plot of the subject mean PCNA LI against the subject mean WCMC.

Fig. 3. Mean percentage of labeling in each quartile, averaged over all subjects, for PCNA and whole-crypt dissection.

Fig. 4. Minimum samples sizes required for 90% power to detect a 25% difference in WCMC between two groups.

Table 1 Comparison of the average proportion of "labeled" cells in each compartment using crypt dissection and PCNA

<table>
<thead>
<tr>
<th>Compartment</th>
<th>Proportion of labeled cells (SD)</th>
<th>Kendall's $\tau$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crypt dissection</td>
<td>PCNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>48.7 (0.035)</td>
<td>53.4 (0.035)</td>
<td>0.10 (0.0792)</td>
</tr>
<tr>
<td>2</td>
<td>46.8 (0.035)</td>
<td>27.7 (0.035)</td>
<td>-0.14 (0.0792)</td>
</tr>
<tr>
<td>3</td>
<td>4.4 (0.035)</td>
<td>5.5 (0.035)</td>
<td>0.14 (0.0792)</td>
</tr>
<tr>
<td>4</td>
<td>0.1 (0.035)</td>
<td>1.7 (0.035)</td>
<td>0.16 (0.0792)</td>
</tr>
</tbody>
</table>

Crypts were divided into four equal compartments by height, starting at the base.

(SD 0.035), and the mean WCMC was 6.58 (SD 3.48). A plot of the subject mean scores for PCNA LI against the subject mean scores for WCMC for these 91 patients is presented in Fig. 2. The plot shows little association between the two indices. Calculation of Kendall's $\tau$ confirms this graphical evidence of little association between the two measures ($\tau = 0.13, P = 0.0792$). Assuming that the subject-to-subject reliability for WCMC is 0.93 and that the subject-to-subject reliability for PCNA LI is 0.99 (16), the correlation between the two measures ($\tau$), corrected for attenuation, is $(0.13^2 / 0.93^2) = 0.14$. Correcting for attenuation does not greatly affect the estimate of the correlation between the two measures because the subject-to-subject reliability was high in both cases.

Compartment Analyses. The overall mean proportions of labeled cells in each quartile for both PCNA and crypt dissection are presented in Table 1. The correlation between the subject average proportion of labeled cells in each compartment obtained using PCNA and crypt dissection ranged from $-0.14$ (quartile 2) to 0.16 (quartile 4). None of the correlations were statistically different from zero. A histogram comparing the overall mean values obtained for each quartile, using both PCNA and crypt dissection, is displayed in Fig. 3. From the figure, it is clear that using PCNA, most of the labeled cells are found near the base of the crypt (first quartile) with much fewer in the second quartile and relatively few in the third and fourth quartiles, at the apex of the crypt. For crypt dissection as well, the mitoses are concentrated in the first and second quartiles with relatively fewer in the upper quartiles. However, for crypt dissection, the proportions in the first and second quartiles were similar on average.

Sample Size Calculations. Fig. 4 displays the results of the sample size calculations for crypt dissection when a difference of 25% between groups is expected. The estimates of the components of variation (and percentage of total variation) due to differences between subjects, differences between biopsies within subjects, and differences between crypts within biopsies were 0.315 (64.8%), 0.030 (6.1%), and 0.142 (29.1%), respectively. From the graph, it is clear that little is gained by reading more than six or seven crypts per biopsy. With 90–100 subjects per group, 95% power could be attained using crypt dissection with just two biopsies per subject and six crypts per biopsy. Based on our previous work (16), for a fixed number of subjects, more biopsies are required for PCNA than for crypt dissection to achieve the same level of power. This is because the percentage of variation due to differences between biopsies is lower for crypt dissection than for PCNA.
Discussion

This study was designed to examine WCMC as a measure of cell proliferation for use in epidemiological and intervention studies in the large bowel. As suggested by others, we found that the crypt dissection technique was simpler to perform than PCNA immunohistochemistry. Cells in mitosis were easy to recognize and count. Examining an entire crypt, rather than a cross-section, requires fewer assumptions about proliferative events within a given crypt.

To accurately portray the biology, a diagnostic test such as crypt dissection must be highly reproducible when repeated on separate occasions. Although reproducibility is generally thought to be straightforward, we discovered some interesting subtleties. Because proliferative measures are generally designed to provide information about a given patient, previous studies have compared the proliferative index, calculated by averaging individual crypt indices. This average, however, is inherently less likely to vary than the individual crypt indices that comprise the average. On repeat, some crypts may have a higher index and some may have a lower index, but the average may be about the same. Strictly speaking, one should compare the index on a crypt-by-crypt basis. Our data suggest that the individual crypt correlations are not as high. This is understandable because the crypt dissection method requires focusing through various planes. A given mitotic figure might be overlooked on repeat. In contrast, previous work with PCNA data has shown the measure to be reliable both in crypt-to-crypt and subject-to-subject comparisons (16).

We also found a poor correlation between LI determined using PCNA immunohistochemistry and WCMC. There are two possible explanations: (a) the two methods measure slightly different events. PCNA measures a broader segment of the cell cycle than crypt dissection, which might partially account for the low correlation; and (b) the second possibility is that the two measures are uncorrelated because WCMC is not a standardized measure. The PCNA LI is standardized in the sense that it represents the percentage of labeled cells in a particular crypt. WCMC measures only the total number of labeled cells. If every crypt contained the same number of cells, this would not be a problem. However, in our data, the total number of cells in a crypt, using PCNA, ranged from 76 to 206 (mean = 124.7; SD = 18.9), while the average number of cells in a crypt for a given subject ranged from 94.7–174.4 (mean = 124.2; SD = 14.8). The total number of cells in a crypt is not the same in every subject or even within a subject. A subject with very large crypts could have a high number of mitoses, but the proportion of cells in mitosis might be low. This inability to correct WCMC for crypt size is a potential problem.

Although it is not possible to adjust the number of mitoses for crypt size, it is possible to determine the location of mitoses within the crypt. Increased proliferation near the crypt apex is thought to be associated with a greater risk of cancer. By determining the proportion of labeled cells and mitoses within discrete quadrants, it might be possible to compare these two measures. However, comparing the subject average proportion of labeled cells appearing in each quadrant for crypt dissection and PCNA, we found no significant correlation between the two measures for any of the quartile comparisons. Furthermore, the distribution of the labeled cells within a crypt appeared to differ for the two methodologies. On average, for PCNA, the greatest concentration of labeled cells was in the first quadrant (crypt base), with smaller proportions of labeled cells in each successive quadrant. For crypt dissection, this trend was not as strong. Most of the labeled cells appeared in the first two quartiles, but the overall mean proportions for the first and second quartiles were roughly the same (48.7% versus 46.8%).

Looking at a number of the drawings made from dissected crypts, we found that in some cases there were no labeled cells at the bottom of the crypt. This migration of the labeled cells in crypt dissection toward the center of the crypt has been noted by previous authors (18). Crypt dissection measures cells further through the mitotic cycle than PCNA, allowing more time for migration. Our analyses confirm that this is often the case and appear to show that PCNA LI and WCMC are not strictly comparable measures of proliferation index.

The sample size estimates for crypt dissection are generally similar to those obtained by others for PCNA (6, 19, 16). If our estimates are slightly higher, this is because we have accounted for the added variation due to within-individual differences in the measures, whereas these differences have not always been accounted for in sample size calculations in the literature. Interestingly, we found the percentage of variation in crypt dissection measures due to differences between biopsies taken on the same subject to be lower than the corresponding percentage for PCNA. Because of this difference, more biopsies would be required for PCNA than for crypt dissection, for a fixed number of subjects, to achieve the same level of power.

In epidemiological studies, it would not be surprising to find differences on the order of 25% between groups. Such a study might require as many as 90 subjects/group. For larger differences, fewer subjects would be necessary. Previous studies using cell proliferation as the outcome measure have demonstrated differences with relatively modest numbers of patients (8, 20). In these studies, the variance estimators used in the test statistics were based on the subject-specific data, ignoring the within-individual variation. Had their variance estimators accounted for the variability between crypts from the same biopsy and between biopsies from the same subject, the results might not have been statistically significant. Alternatively, differences between laboratories and staining procedures which lower the within- and between-subject variation would decrease the sample size needed to show a statistically significant difference between groups.

The results of this study using endoscopic biopsies in humans are at variance with a previous study in rodents (15). Animal studies generally involve giving animals a carcinogen or using a fasting/refeeding strategy. Both of these maneuvers significantly increase proliferation. In relatively healthy, low-risk human subjects, the indices may be sufficiently low to limit the ability to show that the two methods are comparable. The crypt size in litter mates fed a similar diet may also be less variable.

The present study did not assess the utility of PCNA LI and WCMC as correlates of disease risk. The ultimate validation of any marker is its ability to predict disease outcomes (21). The real question is whether any of these measures correlates with risk of developing colorectal cancer or its precursors. The fact that the two measures did not correlate with each other does not imply that they will not be associated with colorectal cancer risk. Alternatively, even if the two measures were highly correlated, they might not be good markers.

The concept of using biomarkers to develop new intervention strategies is an appealing one. Techniques that work in animal models cannot necessarily be directly exported to studies in humans (22). Before biomarkers can be adopted, they must be shown to be reliable. Additional work must be conducted to develop statistical methods that are designed to model these biological variables. The advantages of the method of whole-crypt dissection are considerable, but it remains to be
seen whether this method will be useful to identify groups at risk for colorectal cancer.

In summary, we found poor correlation between the results of whole-crypt dissection and PCNA immunohistochemistry, two techniques that have become popular to measure epithelial cell proliferation. The correlation was poor overall and in comparisons of compartment scores. Additional methodological and statistical research is necessary to determine the reliability, validity, and correlation of the various proliferation measures.

References
Comparison of rectal mucosal proliferation measured by proliferating cell nuclear antigen (PCNA) immunohistochemistry and whole crypt dissection.

S C Murray, R S Sandler, T O Keku, et al.


Updated version  Access the most recent version of this article at:
http://cebp.aacrjournals.org/content/4/7/715

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.