

# Evaluating markers of epithelial-mesenchymal transition to identify cancer patients at risk for metastatic disease

Evan L. Busch<sup>1,2,3</sup> · Temitope O. Keku<sup>4</sup> · David B. Richardson<sup>1</sup> · Stephanie M. Cohen<sup>5</sup> · David A. Eberhard<sup>6</sup> · Christy L. Avery<sup>1</sup> · Robert S. Sandler<sup>1,4</sup>

Received: 3 June 2015 / Accepted: 8 October 2015 / Published online: 27 October 2015  
© Springer Science+Business Media Dordrecht 2015

**Abstract** Most cancer deaths are due to metastases. Markers of epithelial-mesenchymal transition (EMT) measured in primary tumor cancer cells could be helpful to assess patient risk of metastatic disease, even among those otherwise diagnosed with local disease. Previous studies of EMT markers and patient outcomes used inconsistent methods and did not compare the clinical impact of different expression cut points for the same marker. Using digital image analysis, we measured the EMT markers Snail and E-cadherin in primary tumor specimens from 190 subjects in tissue microarrays from a population-based prospective cohort of colorectal cancer patients and

estimated their associations with time-to-death. After measuring continuous marker expression data, we performed a systematic search for the cut point for each marker with the best model fit between dichotomous marker expression and time-to-death. We also assessed the potential clinical impact of different cut points for the same marker. After dichotomizing expression status at the statistically-optimal cut point, we found that Snail expression was not associated with time-to-death. When measured as a weighted average of tumor cores, low E-cadherin expression was associated with a greater risk of dying within 5 years of surgery than high expression (risk difference = 33 %, 95 % confidence interval 3–62 %). Identifying a clinically-optimal cut point for an EMT marker requires trade-offs between strength and precision of the association with patient outcomes, as well as consideration of the number of patients whose treatments might change based on using the marker at a given cut point.

**Electronic supplementary material** The online version of this article (doi:10.1007/s10585-015-9757-7) contains supplementary material, which is available to authorized users.

✉ Robert S. Sandler  
rsandler@med.unc.edu

<sup>1</sup> Department of Epidemiology, University of North Carolina at Chapel Hill, McGavran-Greenberg Hall, CB #7435, Chapel Hill, NC 27599, USA

<sup>2</sup> Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

<sup>3</sup> Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

<sup>4</sup> Department of Medicine, University of North Carolina at Chapel Hill, 4157 Bioinformatics Building, CB #7555, Chapel Hill, NC 27599, USA

<sup>5</sup> Translational Pathology Laboratory, University of North Carolina at Chapel Hill, 919 Brinkhous-Bullitt, CB #7525, Chapel Hill, NC 27599, USA

<sup>6</sup> Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, 450 West Drive, CB #7295, Chapel Hill, NC 27599, USA

**Keywords** Metastasis · Epithelial-mesenchymal transition · Outcomes · Diagnosis · Risk · Stratification

## Introduction

Roughly 90 % of cancer deaths are due to metastases [1]. Current methods to assess metastatic disease at diagnosis—examination of regional lymph nodes and radiologic imaging—do not always successfully detect metastases that are present, leading to false diagnoses of local disease. For example, approximately 25 % of colorectal cancer (CRC) patients diagnosed with local disease later experience recurrence [2]. Markers of epithelial-mesenchymal transition (EMT) [3–5] measured in primary tumor cancer cells could provide an additional test of risk for metastatic

disease to guide treatment decisions, even when metastases are not detected by the two standard tests.

EMT is a mechanism of cancer cell metastasis that involves epithelial cells temporarily becoming mesenchymal cells [4]. This occurs when cellular expression levels of EMT inducers increase, triggering decreased expression of epithelial markers and increased expression of mesenchymal markers. These changes lead to loss of adhesion to adjacent cells—enabling detachment of the transitioning cell from the primary tumor—as well as modifications that enhance motility and invasiveness.

Past studies of associations between EMT markers in primary tumors and patient outcomes have measured and analyzed marker expression data using inconsistent methods [6]. For instance, different studies of the same marker measured using the same laboratory technique have often used different scoring scales and defined high versus low expression differently. Ordinal marker expression data have frequently been collected, limiting the ability to identify clinically useful cut points to define marker expression status. The lack of uniform, clinically-informative methods across studies could contribute to inconsistent findings and hamper translation of EMT markers to clinical use.

To facilitate translation of EMT markers, we developed new approaches that permit direct evaluation of how an EMT marker could be implemented clinically. We focused on digital image analysis of immunostained tissue specimens to obtain continuous marker expression data, statistical evaluation of the clinical utility of a given cut point to define high versus low marker expression, and rigorous selection of covariates to include in multivariate models.

To illustrate our approach, we measured the EMT inducer Snail and epithelial marker E-cadherin in primary tumors from a population-based prospective cohort study of CRC mortality and estimated their associations with time-to-death. We hypothesized that low expression of E-cadherin and high expression of Snail would be associated with shorter times from surgery to death compared to opposite expression levels.

## Materials and methods

### Study population

Subjects were enrolled in the Cancer Care Outcomes Research and Surveillance Consortium (CanCORS), a population-based, prospective, case-only, multi-site observational study of colorectal and lung cancer patients [7]. Briefly, the study assessed the impact of health-system, provider, and patient factors on cancer outcomes. Patients

were at least 21 years of age at diagnosis and were enrolled within 3 months of diagnosis during 2003–2006. The study collected patient surveys, surrogate surveys for patients who were deceased or too ill to participate, and medical records data [8]. Vital status for all subjects was verified using the Social Security Death Index on 4 May 2010, providing at least 42 months of follow-up observation per individual.

The North Carolina CanCORS site enrolled 990 CRC patients and was the only one to collect tumor specimens. Subjects in the present biomarker study came from a catchment area of 33 counties in eastern and central North Carolina at the time of diagnosis. Investigators obtained primary tumor and non-neoplastic adjacent colorectal tissue samples from 506 subjects.

Formalin-fixed, paraffin-embedded tissue specimens were sent from hospitals across the catchment area to the University of North Carolina at Chapel Hill (UNC), where they were used to construct tissue microarrays (TMAs) as described previously [9]. Most subjects had multiple cores from both primary tumor and “normal” (non-neoplastic) margin. We measured EMT markers in 12 representative TMAs that included specimens from 219 subjects. To be included in the study sample, subjects had to have at least one core of tumor tissue successfully stained for one marker, with that core having at least 50 epithelial cells and unambiguous histology. From the 12 TMAs, we excluded 26 subjects lacking adequate tumor tissue and an additional 3 subjects who could not be linked to medical records data, yielding a final study sample of 190 subjects.

### Immunohistochemistry

We selected EMT markers for evaluation based on the results of previous studies and criteria discussed in our prior literature review [6]. Marker protein expression was measured using the following antibodies: E-cadherin [mouse monoclonal ready to use (RTU), clone 36B5 (cat #PA0387) from Leica Microsystems Inc. (Norwell, MA)] and Snail [goat polyclonal (ab53519) from Abcam (Cambridge, MA)].

Immunohistochemistry (IHC) was performed at the UNC Translational Pathology Laboratory (TPL) using the Bond fully-automated slide staining system (Leica Microsystems Inc., Norwell, MA). Slides were deparaffinized in Bond Dewax solution (AR9222) and hydrated in Bond Wash solution (AR9590). Antigen retrieval was performed at 100 °C for Snail (for 30 min) in Bond-epitope retrieval solution 1 at pH 6.0 (AR9961) and for E-cadherin for 20 min at 100 °C in solution 2 at pH 9.0 (AR9640). After pretreatment, anti-E-cadherin was applied for 15 min and anti-Snail (1:200) for 30 min.

Detection of Snail was performed using the Bond Intense R Detection System (DS9263) supplemented with the LSAB + kit (DAKO, Carpinteria, CA). E-cadherin detection used the Bond Polymer Refine Detection System (DS9800). Stained slides were dehydrated and cover-slipped. Positive and negative controls (no primary antibody) were included for each antibody. All assays were single-marker (i.e., no multiplex assays).

To verify E-cadherin antibody specificity, tonsil tissue was stained both with and without primary antibody (positive and negative controls, respectively). Snail antibody specificity was confirmed in normal kidney (positive control) and normal liver (negative control) [10, 11]. Additional negative controls were performed with goat IgG (Santa Cruz, sc-2755) used in place of the Snail antibody. Both antibodies stained appropriately in the relevant sub-cellular compartments (E-cadherin, plasma membrane; Snail, nucleus) in each tissue (see Online Resource 3, Figs. S1 and S2).

Stained slides were digitally imaged at  $\times 20$  magnification using the Aperio ScanScope XT (Aperio Technologies, Vista, CA). Digital images were stored in the Aperio Spectrum Database.

Example images of staining for colorectal tissue from CanCORS subjects are provided in Online Resource 3 (Fig. S3).

### Automated analysis of digital IHC images

Computer algorithms annotated and scored every eligible tissue core to obtain continuous marker expression data. We used approximately 65 cores originating from two TMAs for algorithm training and automated-analysis validation.

Definiens Composer Technology (Tissue Studio version 2.1.1 with Tissue Studio Library version 3.6.1; Definiens Inc., Carlsbad, CA) was used to annotate images for regions enriched in epithelial cells in IHC-stained TMA cores. To detect differences in cell shape and tissue structure, we developed two Composer algorithms per marker—for non-neoplastic adjacent and tumor tissue, respectively—as both types of tissue were present on each TMA.

We then developed two Tissue Studio scoring algorithms (“solutions”) per marker (non-neoplastic and tumor tissue). E-cadherin membrane and cytoplasmic staining were measured on a continuous average intensity scale of 0–3. Snail was measured as core percent positive nuclei on a continuous scale of 0–100. Several additional Snail scoring algorithms were developed, but only scores based on the first algorithm were used for modeling (see Online Resource 1 for further details).

To evaluate the reliability of computer annotations, one of us (ELB) used Aperio ImageScope (version 11.2; Leica Biosystem, Buffalo Grove, IL) to manually annotate the same 65 cores per marker that were used to optimize Tissue Studio solutions. He remained blind to patient and tumor characteristics while annotating. Automated scores obtained via manual and automated annotation produced Pearson correlations of 0.91 for E-cadherin and 0.94 for Snail. All 12 TMAs stained for E-cadherin and Snail were then analyzed (24 slides in total).

Subjects typically had multiple cores available of a given tissue type (tumor or non-neoplastic). To assign an expression value for each subject by marker and tissue type, we handled replicate cores in two ways: first, as a weighted average of cores, and second, by assigning the expression value of the subject’s “worst core” as the marker expression value. For weighted averages, the weights were area analyzed for E-cadherin and number of nuclei for Snail. The worst core by tissue type was assigned as the lowest core average intensity for E-cadherin and as the highest core percent positive nuclei for Snail.

See Online Resource 1 for further details on algorithm development and validation. See Online Resource 3 for example images of annotations and scoring in colorectal tumor and non-neoplastic tissue (Figs. S4, S5).

### Outcome

In statistical models, the dependent variable was length of time in days from primary tumor surgery until all-cause mortality, with administrative censoring at 5 years after surgery.

### Covariates

Covariates for multivariate statistical models were selected based on prior studies [6], considerations of biological plausibility, and directed acyclic graph theory [12].

Per directed acyclic graph theory, we adjusted for those variables considered to be common causes of both the independent variable of interest (EMT marker expression in primary tumor cancer cells) and the dependent variable (time from surgery to death): age, neoadjuvant chemotherapy, neoadjuvant radiation therapy, tumor size (T-stage), lymph-node metastasis diagnosis (N-stage), and distant-metastasis diagnosis (M-stage). We used overall TNM stage as a single covariate instead of adjusting for the component stages as three separate variables since including both overall stage and any of the component stages would constitute inappropriate overadjustment.

Pre-surgery treatments consisted of neoadjuvant chemotherapy and neoadjuvant radiation therapy as two

separate variables, each coded as received or not received. Cancer treatments (chemotherapy and radiation therapy) prior to surgery were included as covariates but treatments administered after surgery were not included. While treatments at either time are related to time-to-death, the observed EMT marker expression in primary tumor cancer cells could only have been exposed to treatments before surgery because the primary tumor remained in the body at that point, whereas the tumor was removed before the administration of post-surgery treatments. Based on directed acyclic graph theory, we concluded that other variables related to CRC prognosis—such as microsatellite status, tumor budding, and KRAS status—should not be adjusted for in multivariate models.

### Identification of statistically-optimal marker expression cut point

For each continuous marker expression variable, we identified the cut point distinguishing high expression from low expression that was most strongly associated with time-to-death.

For every possible cut point along any marker expression continuum, we defined high expression as expression at or above the cut point and low expression as expression below the cut point. Whether high expression is clinically desirable depends on the particular marker. High E-cadherin status would be expected to correlate with better outcomes (i.e., longer time-to-death) [4, 6, 13], whereas high Snail [4, 6] status would be expected to correlate with worse outcomes.

To identify the statistically-optimal cut point for each continuous marker expression variable, we iteratively dichotomized marker expression at every possible cut point in the observed tumor tissue data, with each cut point corresponding to a different subject's expression value. Each dichotomization of marker expression status was fit as the only independent variable in a Cox proportional hazards model with time-to-death as the outcome, producing a model fit statistic. The cut point with the lowest model fit statistic was considered statistically optimal.

For macro SAS code and further details, see Online Resource 2.

### Statistical analysis

The data collection yielded four continuous marker expression variables: weighted average and worst core for each of the two markers. We first used unpaired two-sample t-tests to assess whether average continuous marker expression differed between tumor and non-neoplastic tissue. All subsequent analyses used tumor tissue only. We

applied the macro to the tumor tissue data for the four continuous marker expression variables to identify the statistically-optimal cut point for each. Every optimal cut point was used to create a dichotomous marker expression variable (low versus high for E-cadherin, high versus low for Snail).

We generated Kaplan–Meier survival curves stratified by dichotomous marker expression status for one marker or two markers jointly, assessing differences between strata using the logrank test. Next, for each optimally-dichotomous marker expression variable, we fit unadjusted and adjusted Cox proportional hazards models of time-to-death. For marker expression variables found to be associated with patient outcomes at their statistically-optimal cut points, we evaluated additional cut points to explore whether the statistically-optimal cut point ought to be considered the clinically-optimal cut point.

Prior to modeling, non-informative observations (e.g. “No Answer,” “Don't Know,” “Unknown”) were recoded as missing. Missing data for all model variables were evaluated using multiple imputation. P values of 0.05 or below were considered statistically significant. All analyses were performed using SAS 9.3 (SAS Institute, Cary, NC). The Institutional Review Board at UNC approved the protocol. All subjects provided informed consent.

### Results

Subject characteristics are shown in Table 1. We found no differences between overall North Carolina CanCORS and the subset for whom EMT markers were measured in primary tumors.

On average, tumor tissue had lower E-cadherin expression than non-neoplastic adjacent tissue regardless of whether expression values were assigned as a weighted average of cores or as the worst core (Table 2). However, average Snail expression was higher in non-neoplastic tissue than in tumor tissue (Table 2). While the difference was not large, this relationship was consistent across both ways of assigning expression values and all three Snail scoring algorithms (Table S1 in Online Resource 3).

On the average intensity scale of 0–3 for E-cadherin, we found that the statistically-optimal cut point was about 0.52 for weighted averages and 0.42 for worst cores. On a percent positive nuclei scale for Snail using the first nuclear scoring algorithm, the statistically-optimal cut point was about 25.2 % for weighted averages and 63.6 % for worst cores. For each marker expression variable, Table 3 presents the number of subjects by statistically-optimal dichotomous marker expression status.

For E-cadherin weighted averages, subjects with low tumor expression had worse survival than those with high

**Table 1** Subject characteristics for overall North Carolina CanCORS and subset in whose primary tumors EMT markers were measured

Characteristic	mean (SD) or N (%) <sup>a</sup>	Overall NC-CanCORS		EMT study sample		P value
		(N = 990)		(N = 190)		
Age at baseline (years)	66.3	13.2	66.6	13.1	0.8	
Sex					0.6	
Male	488	49.3 %	90	47.4 %		
Female	502	50.7 %	100	52.6 %		
Race					0.6	
Non-hispanic white	762	77.0 %	150	78.9 %		
Other	228	23.0 %	40	21.1 %		
Tumor stage					0.9	
Local	437	51.1 %	100	52.6 %		
Regional	311	36.4 %	66	34.7 %		
Distant	107	12.5 %	24	12.6 %		
Missing/noninformative	135		0			
Tumor location					0.07	
Proximal	279	42.9 %	82	50.9 %		
Distal	371	57.1 %	79	49.1 %		
Missing/noninformative	340		29			
Tumor grade					0.4	
Well/moderately differentiated	539	83.7 %	129	81.1 %		
Poorly/not differentiated	105	16.3 %	30	18.9 %		
Missing/noninformative	346		31			
Neoadjuvant chemotherapy					0.3	
Yes	46	8.5 %	8	5.7 %		
No	496	91.5 %	132	94.3 %		
Missing/noninformative	448		50			
Neoadjuvant radiation					0.09	
Yes	41	7.6 %	5	3.6 %		
No	501	92.4 %	135	96.4 %		
Missing/noninformative	448		50			
Died within 5 years of baseline	366	37.0 %	62	32.6 %	0.3	

<sup>a</sup> Percentages are for informative non-missing data

**Table 2** Average continuous EMT marker expression in tumor tissue compared to non-neoplastic adjacent tissue

Marker	Algorithm	Scale	Tissue type	N	Marker expression					
					Weighted average <sup>a</sup>			Worst core <sup>b</sup>		
					Mean	SD	P value	Mean	SD	P value
E-cadherin	Membrane + cytoplasm	Average intensity (0–3)	Tumor	188	0.84	0.22	<0.0001	0.74	0.24	<0.0001
			Non-neoplastic	181	0.98	0.22		0.87	0.28	
Snail	Nuclear <sup>c</sup>	Percent positive nuclei (0–100)	Tumor	185	41.7	22.1	0.02	52.3	24.9	0.0004
			Non-neoplastic	173	47.9	27.9		62.1	27.2	

<sup>a</sup> Marker expression values assigned as weighted average of cores by tissue type (weighted by area analyzed for E-cadherin and number of nuclei for Snail)

<sup>b</sup> Marker expression values assigned as expression by tissue type of the core with lowest expression for E-cadherin and highest expression for Snail

<sup>c</sup> First nuclear Snail algorithm of the multiple Snail scoring algorithms that were developed

**Table 3** Statistically-optimal dichotomous EMT marker expression status cross-tabulated with tumor stage and with risk of dying within 5 years of surgery

	E-cadherin				Snail			
	Weighted Average <sup>a</sup>		Worst Core <sup>b</sup>		Weighted Average <sup>a</sup>		Worst Core <sup>b</sup>	
	Low	High	Low	High	Low	High	Low	High
Stage								
Local	6	93	3	96	29	69	54	44
Regional	5	61	3	63	14	51	45	20
Distant	0	23	0	23	1	21	13	9
Mortality risk								
Died within 5 years (x)	7	55	3	59	17	44	43	18
Total at risk (y)	11	177	6	182	44	141	112	73
5-Year risk of death post-surgery (%) (x/y)	64	31	50	32	39	31	38	25

Of 190 overall subjects in the EMT study, the number with tumor tissue marker data available was 188 for E-cadherin and 185 for Snail

Low expression status is expression in tumor tissue below the statistically-optimal cut point and high expression status is expression in tumor tissue at or above the statistically-optimal cut point. The a priori hypotheses were that low E-cadherin subjects would have worse outcomes while high Snail subjects would have worse outcomes. For the value of the statistically-optimal cut point for each marker expression variable, see “Materials and methods” section

<sup>a</sup> Marker expression values assigned as weighted average of cores in tumor tissue (weighted by area analyzed for E-cadherin and number of nuclei for Snail)

<sup>b</sup> Marker expression values assigned as expression in tumor tissue of the core with lowest expression for E-cadherin and highest expression for Snail

tumor expression (Fig. 1). None of the Kaplan–Meier curves stratified by the other three optimally-dichotomous marker expression variables revealed a statistically-significant difference in survival (not shown). A survival curve jointly stratified by E-cadherin and Snail weighted average status had a significant logrank test (see Online Resource 3, Fig. S6). Among the strata of joint expression status, no pattern consistent with the study hypotheses was observed.

Unadjusted and adjusted proportional hazards model results paralleled the single-variable stratified survival curves. Low E-cadherin weighted average expression was associated with greater hazards of dying than high expression both when unadjusted [Hazard ratio (HR) = 2.84, 95 % Confidence Interval (CI) 1.29–6.28] and adjusted (HR = 2.57, 95 % CI 1.10–6.03), while no associations were found for any of the other optimally-dichotomous expression variables (Table 4).

To evaluate whether the statistically-optimal cut point would be clinically optimal, we explored several trade-offs between strength of cut-point/time-to-death association and the number of subjects whose treatments might change due to clinical use of EMT markers. Specifically, we considered three different E-cadherin weighted average cut points that were either statistically significant or nearly so: about 0.52 (statistically-optimal value), 0.60, and 0.85 (Table 5). Setting the cut point to a value other than the statistically-optimal value led to hazard ratio point estimates that were weaker than the one at the optimal cut point, but marker expression status was still effectively associated with

outcomes at each of these cut points. Notably, the precision of hazard ratio estimates was better at cut points other than the statistically-optimal value, with a confidence limit ratio of 5.48 at a cut point of 0.52, 3.48 at a cut point of 0.60, and 3.11 at a cut point of 0.85.

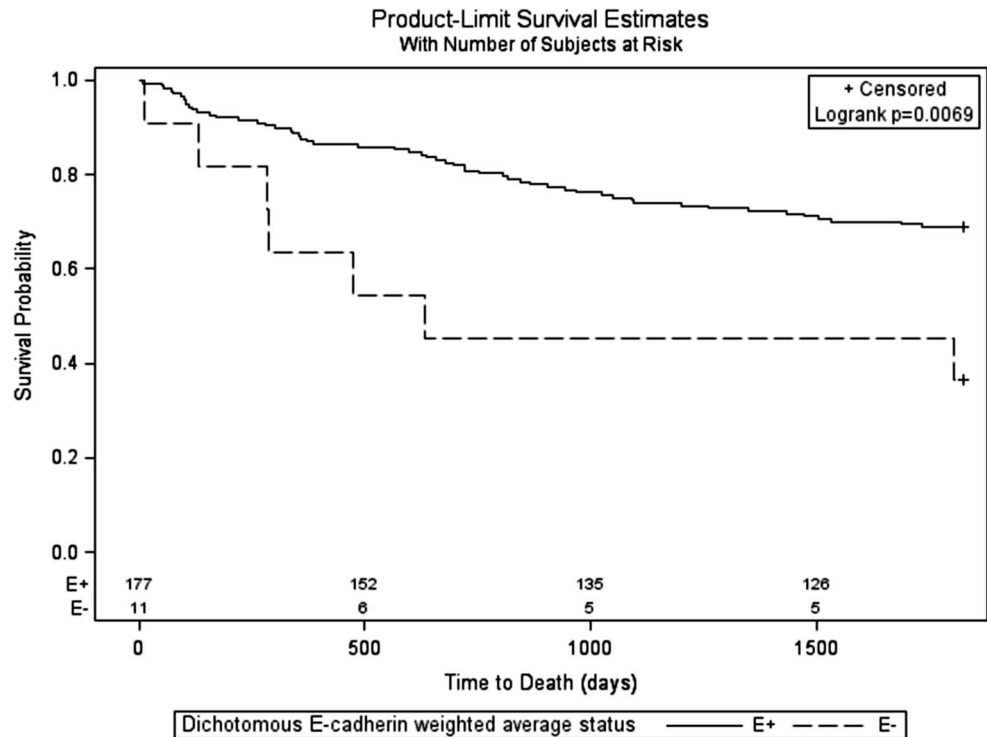
The number of subjects whose treatments might change based on E-cadherin measurements—those diagnosed with local disease but who had low E-cadherin expression—varied substantially with cut point. Of 99 subjects with E-cadherin measurements and diagnosed with local disease, 6 had low expression at the statistically-optimal cut point, 16 at a cut point of 0.60, and 56 at a cut point of 0.85 (Table 5).

## Discussion

EMT markers measured in primary tumor cancer cells could potentially improve the accuracy of cancer staging by identifying patients at risk for metastatic disease who have false negative test results for both lymph node evaluation and radiologic imaging. This improvement in staging could lead to more appropriate treatment decisions that would reduce the number of patients diagnosed with local disease who later experience recurrence, ultimately improving survival outcomes.

Successful translation requires consistent, clinically-informative design of studies of EMT markers and patient outcomes. Although marker expression is naturally

**Fig. 1** Kaplan-Meier overall survival stratified by expression status of E-cadherin measured as a continuous weighted average of tumor cores and then dichotomized at the statistically-optimal cut point (E+ = high expression, E- = low expression)



**Table 4** Unadjusted and adjusted Cox proportional hazards models of the effect of optimally-dichotomized marker expression status on time-to-death censored at 5 years after surgery (n = 190)

Variable	Weighted average <sup>a</sup>				Worst core <sup>b</sup>			
	Unadjusted		Adjusted <sup>c</sup>		Unadjusted		Adjusted <sup>c</sup>	
	HR	95 % CI	HR	95 % CI	HR	95 % CI	HR	95 % CI
E-cadherin <sup>d</sup>	2.84	1.29, 6.28	2.57	1.10, 6.03	1.53	0.47, 4.95	1.56	0.32, 7.63
Snail <sup>e</sup>	0.83	0.48, 1.45	0.61	0.33, 1.13	0.62	0.36, 1.08	0.57	0.32, 1.02

Prior to dichotomization, continuous E-cadherin measured as core average intensity (0–3) and continuous Snail measured as core percent positive nuclei (Snail) (0–100). Subjects with missing data for a given marker expression variable had values imputed to retain full sample size. For the value of the statistically-optimal cut point for each marker expression variable, see “Materials and methods” section

<sup>a</sup> Marker expression values assigned as weighted average of tumor cores (weighted by area analyzed for E-cadherin and number of nuclei for Snail)

<sup>b</sup> Marker expression values assigned as average intensity of the subject’s tumor core with the lowest average intensity for E-cadherin, or as percent positive nuclei of the subject’s tumor core with the highest percent positive nuclei for Snail

<sup>c</sup> Adjusted for age (continuous), TNM stage (local/regional/distant), neoadjuvant chemotherapy (yes/no), and neoadjuvant radiation therapy (yes/no). Expression status for a given marker was not adjusted for the other markers

<sup>d</sup> Comparison is low E-cadherin expression to high E-cadherin expression

<sup>e</sup> Comparison is high Snail expression to low Snail expression

continuous, the clinical purpose of the markers is to guide therapy. Since treatment decisions are inherently binary, what matters is dichotomous marker expression. Therefore, the choice of cut point to dichotomize continuous marker expression is critical.

We developed approaches to EMT marker measurement and analysis that allow evaluation of the clinical utility of

an EMT marker at different cut points. As a proof of principle, we measured two EMT markers in a set of CRC primary tumors from a population-based prospective cohort study. We found that E-cadherin expression measured as a weighted average of tumor cores was associated with time to all-cause mortality independent of tumor stage, but Snail expression was not associated with outcomes. This implied

**Table 5** Trade-offs between strength and precision of cut-point/time-to-death association and number of patients with low or high expression, by stage distribution and 5-year risk of death, for multiple cut points of E-cadherin weighted average expression

	Cut point 0.52 <sup>b</sup>		Cut point 0.60		Cut point 0.85	
	HR	95 % CI (CLR)	HR	95 % CI (CLR)	HR	95 % CI (CLR)
Cox model estimate <sup>a</sup>	2.57	1.10, 6.03 (5.48)	2.40	1.29, 4.49 (3.48)	1.75	0.99, 3.08 (3.11)
Stage	Low	High	Low	High	Low	High
Local (n = 99)	6	93	16	83	56	43
Regional (n = 66)	5	61	11	55	41	25
Distant (n = 23)	0	23	1	22	11	12
Mortality risk						
Died within 5 years (n = 62) (x)	7	55	14	48	42	20
Total at risk (n = 188) (y)	11	177	28	160	108	80
5-year risk of death post-surgery (%) (x/y)	64	31	50	30	39	25
Mortality Risk Effect estimates <sup>c</sup>	Estimate	95 % CI	Estimate	95 % CI	Estimate	95 % CI
Risk difference (%)	33	3, 62	20	0, 40	14	1, 27
Risk ratio	2.05	1.25, 3.37	1.67	1.07, 2.59	1.56	0.99, 2.43

CLR Confidence limit ratio (upper limit/lower limit)

E-cadherin measured as a weighted average (weighted by area analyzed) of continuous tumor core average intensities (0–3) prior to dichotomization. Of 190 subjects, 2 had missing data for E-cadherin, but multiple imputation enabled retention of these 2 in Cox models

<sup>a</sup> For effect of dichotomous E-cadherin expression (low versus high) on time-to-death, adjusted for age (continuous), TNM stage (local/regional/distant), neoadjuvant chemotherapy (yes/no), and neoadjuvant radiation treatments (yes/no)

<sup>b</sup> Statistically-optimal cut point by best model fit

<sup>c</sup> Unadjusted comparisons of low versus high expression status

that E-cadherin has promise as a marker to identify CRC patients at risk for metastatic disease independent of lymph node evaluation and imaging results.

Our results suggested that at least three criteria should be used to evaluate the association between dichotomous EMT marker expression at a given cut point and patient outcomes: strength of point estimate, precision, and the number of patients whose treatments would be changed by implementing the marker clinically at that cut point. Table 5 illustrates the information needed to compare the performance of an EMT marker at different cut points.

The table shows that, while multiple E-cadherin cut points were associated with time to all-cause mortality, there were trade-offs between them in terms of clinical performance. On a continuous average intensity scale of 0–3, the statistically-optimal cut point (in terms of model fit) of 0.52 had the strongest point estimate in terms of being furthest from the null value. A cut point of 0.85 had the best precision.

In terms of the patients whose treatments would change by implementing E-cadherin as a diagnostic tool, this would consist of those diagnosed with local disease according to lymph node evaluation and imaging but who had low E-cadherin expression. Without knowledge of E-cadherin status, these patients would not generally

receive adjuvant chemotherapy [14], but after introducing the marker, they might be considered for chemotherapy. Recall that about 25 % of CRC patients diagnosed with local disease later experience recurrence [2]. This suggests that a useful cut point for E-cadherin in CRC would yield an approximate distribution of 25 % low expression and 75 % high expression among those diagnosed with local disease according to lymph node evaluation and imaging. Our study had 99 subjects with E-cadherin measurements who were diagnosed with local disease. The recurrence data for CRC patients diagnosed with local disease implies that an effective cut point would yield a distribution of about 25 low E-cadherin and 74 high E-cadherin expression status among these 99 subjects. Of the three cut points in Table 5, a cut point of 0.60 came closest to this distribution.

Our approach had several important strengths. First, digital image analysis yielding continuous marker expression data maximized the opportunity to identify the most clinically useful cut point [15]. In contrast, collecting dichotomous or ordinal marker data, such as are obtained by manual IHC scoring, would obscure much of the natural variation that may be present and greatly reduces the number of cut points that can be explored. Second, our approach to analyzing the data permitted direct evaluation



of the clinical impact of using a particular marker at a given cut point. It also allowed straightforward comparison of the strengths and weaknesses of different cut points for the same marker in the same study sample. Choosing cut points based on biologically arbitrary criteria such as percentiles of marker expression is not recommended [16].

Third, the selection of covariates for multivariate models has varied considerably and been without justification in previous studies [6]. Our approach was to identify a wide range of possibilities using prior literature and then determine whether to include or exclude each possibility based on directed acyclic graph theory [12]. Finally, our sample was population-based, implying that our results could have greater generalizability than the hospital-based samples that have dominated prior studies [6, 17]. Although the subjects in this analysis were approximately 20 % of those enrolled in North Carolina CanCORS, we found no differences in subject characteristics between overall North Carolina CanCORS and those included in the EMT study sample (Table 1). This suggested that the EMT study sample remained representative of the underlying source population.

In terms of limitations, first, the only available outcome was time from surgery to all-cause mortality. It would have been informative to also examine associations between marker expression and time to cancer-specific mortality as well as time to recurrence, but these alternative outcomes were not measured in CanCORS. Second, the sample size was small, which prevented division of the subjects into separate training and validation sets with sufficient power to detect an association between marker expression and outcomes.

Finally, our study did not sample tumors in a consistent way. Ideally, every tumor would have been sampled at each of the invasive front, tumor center, and an edge of the tumor away from the invasive front [6]. However, for any given tumor in our dataset, we did not know from which part of the tumor our tissue specimens came. EMT marker expression could vary throughout a tumor and it may be that, for clinical purposes, physicians should always sample a particular portion (e.g. the invasive front). Had systematic tumor sampling been performed, we could have calculated portion-specific estimates of, say, the association between invasive front E-cadherin and time-to-death, and separately, the association between tumor center E-cadherin and time-to-death. Unfortunately, we could not analyze the data this way given the random tumor sampling.

An important implication of this study for clinical cancer pathology is that EMT marker immunohistochemical measurements should be incorporated into clinical practice using digital image analysis to obtain a continuous expression score on a scale standardized for the particular marker. This score should then be compared to a cut point

specific to the marker that has been identified via epidemiologic studies and is based on a consideration of the criteria discussed above: strength of association with patient outcomes, precision of association, and proportion of patients diagnosed with local disease according to lymph node evaluation and radiologic imaging whose treatments would change based on EMT marker status. For CRC, our results suggest that a finding of low E-cadherin levels could signal to clinicians that cancer cells were likely detaching from the primary tumor prior to surgery.

This study furthers translational EMT research in both specific and general ways. Specifically, our results imply that E-cadherin is a promising marker to identify CRC patients at risk for metastatic disease even when metastases are not detected by lymph node evaluation and imaging. More generally, we have developed an approach to measurement and analysis of EMT marker data that is more rigorous, thorough, and clinically-informative than techniques described in previous studies. This approach is not specific to CRC and could benefit future studies in different tumor sites. Adoption of our approach could facilitate translation of EMT markers to clinical use and thereby improve outcomes across a wide spectrum of cancer patients.

**Acknowledgments** This research was funded in part by Grants from the National Institutes of Health (P30 DK034987, U01 CA93326). The Translational Pathology Laboratory is supported in part by grants from the National Cancer Institute (3P30CA016086) and the University of North Carolina at Chapel Hill University Cancer Research Fund. ELB was supported in part by a Grant from the National Cancer Institute (5T32CA009001).

#### Compliance with ethical standards

**Conflicts of interest** The authors have no conflicts of interest to report.

**Human and animal rights and informed consent** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

## References

- Weinberg RA (2007) The biology of cancer. Garland Science, New York
- Young PE et al (2014) Early detection of colorectal cancer recurrence in patients undergoing surgery with curative intent: current status and challenges. *J Cancer* 5(4):262–271
- Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144(5):646–674
- Kalluri R, Weinberg RA (2009) The basics of epithelial-mesenchymal transition. *J Clin Invest* 119(6):1420–1428
- Thiery JP et al (2009) Epithelial-mesenchymal transitions in development and disease. *Cell* 139(5):871–890

6. Busch EL, McGraw KA, Sandler RS (2014) The potential for markers of epithelial-mesenchymal transition to improve colorectal cancer outcomes: a systematic review. *Cancer Epidemiol Biomark Prev* 23(7):1164–1175
7. Ayanian JZ et al (2004) Understanding cancer treatment and outcomes: the Cancer Care Outcomes Research and Surveillance Consortium. *J Clin Oncol* 22(15):2992–2996
8. Malin JL et al (2006) Understanding cancer patients' experience and outcomes: development and pilot study of the Cancer Care Outcomes Research and Surveillance patient survey. *Support Care Cancer* 14(8):837–848
9. Kang M et al (2013) Somatic gene mutations in African Americans may predict worse outcomes in colorectal cancer. *Cancer Biomark* 13(5):359–366
10. Berglund L et al (2008) A genecentric human protein atlas for expression profiles based on antibodies. *Mol Cell Proteomics* 7(10):2019–2027
11. Yao D, Peng S, Dai C (2013) The role of hepatocyte nuclear factor 4alpha in metastatic tumor formation of hepatocellular carcinoma and its close relationship with the mesenchymal-epithelial transition markers. *BMC Cancer* 13:432
12. Greenland S, Pearl J, Robins JM (1999) Causal diagrams for epidemiologic research. *Epidemiology (Cambridge, Mass)* 10(1):37–48
13. He X et al (2013) Downregulated E-cadherin expression indicates worse prognosis in Asian patients with colorectal cancer: evidence from meta-analysis. *PLoS One* 8(7):e70858
14. Benson AB 3rd et al (2004) American Society of Clinical Oncology recommendations on adjuvant chemotherapy for stage II colon cancer. *J Clin Oncol* 22(16):3408–3419
15. Rizzardi AE et al (2012) Quantitative comparison of immunohistochemical staining measured by digital image analysis versus pathologist visual scoring. *Diagn Pathol* 7:42
16. Rothman KJ (2014) Six persistent research misconceptions. *J Gen Intern Med* 29(7):1060–1064
17. Szklo M (1998) Population-based cohort studies. *Epidemiol Rev* 20(1):81–90