Optimal use of staging data in international comparisons of colorectal cancer survival

Dear Editor,

We welcome the comments by Eden et al. (1), which provide an opportunity to discuss the complex issue of international comparisons of survival by stage at diagnosis (2). It is simply not the case that we compared exclusively pathological data on Dukes' stage from the UK with optimally integrated data on stage from other countries, as Eden et al. suggest. Eden et al. assert that we "have principally misinterpreted the origins and composition of the data used in [the] analysis". We strongly disagree. An expert committee of cancer registry directors and senior colorectal oncologists in each country was actively involved at every stage of this study, from protocol design to data quality control, analysis, interpretation and drafting, together with us, under the direction of an international Programme Board.

We were aware of the caveats highlighted by Eden et al., and we took great care to examine the pathological (p) and clinical (c) origins of the stage data for every patient in these data sets. We defined an algorithm such that Dukes' stage was determined via the component pT, pN and cM values, wherever possible. We also noted the possible biases and differences that can arise if only the 'grouped' Dukes' stage was available. This is described in the 'methods paper' that accompanied our analyses (3). We acknowledged in that paper that some inconsistencies remained, and we reiterated it in the colorectal cancer paper (2). It is precisely because of the "limitations of this methodology" that we presented two sets of analyses, one based on SEER Summary Stage 2000 and another based on Dukes' stage.

Table 1 below shows the origin of the final stage data used in the survival analyses. Among patients with a valid stage (defined in (3)), we give the proportions of patients whose final stage was specified on the basis of component T, N and M codes supplied by the registry. Among those, we show the origin of those variables as either pathological, clinical or unknown. If one of the component codes is shown as of 'unknown' origin, it means the registry did not indicate whether the value provided was based on pathological or clinical evidence: it may thus reflect *either* a value that had been deliberately integrated by the source registry as the best reflection of the stage on the basis of all available data, *or* that the pathological or clinical origin of the value was actually unknown. It was not possible to distinguish between these scenarios from the available data or from discussion with the registries.

The proportion of rectal cancer patients for whom component T, N and M codes were supplied was 0% in Norway. In Sweden, the proportion was just over 30% for both colon and rectal cancer, compared to just under 20% in the UK – not a great difference. Among those patients for whom T, N and M codes *were* available, the UK had relatively low proportions of pT, pN and cM codes (the optimal values). Where T, N and M codes were available for patients with colon or rectal cancer, their origin was unknown for about one-third (27-35%) of patients in the UK, but this proportion reached 90-100% of patients in Denmark and Sweden.

Eden et al.'s description of how stage was defined in the ICBP study is simplistic. They assume that we based our final stage variable on pathological Dukes' stage alone, although as shown in Table 1, the picture is more complex. Wherever possible, we used much more detailed information.

Eden et al. exemplify their doubts with a combined analysis of patients diagnosed with either colon or rectal cancer in 2011, in one region of England with good data on stage, ignoring patients with missing stage data. Those patients were diagnosed 4-11 years later than the patients in this ICBP study (2000-2007), which covered most of England. Moreover, the "ICBP-like" distribution of stage in their Figure 1 is not even close to the distribution we published (e.g. 2% metastatic, cf. 16.9% in the ICBP study). It is not an accurate description of the data we analysed.

Their distribution of "ICBP-like stage" for colon and rectal cancers combined is closer to the distribution in the ICBP study for those patients whose stage was based solely on a Dukes' stage variable (Table 2). The relatively low proportion of metastatic disease among these patients because of pathological Dukes' staging was noted very early in the ICBP study: it was the subject of correspondence with scientists in the English cancer registries, including Dr Rous. When patients whose final stage was constructed from the component T, N and M codes are included, the overall proportion with metastatic disease in the ICBP study (17%) is similar to the proportion in Eden et al.'s analysis using integrated stage (14%).

We agree that "international comparisons must understand the source of the data". The ICBP studies have confirmed the need for better communication between clinicians and cancer registries. Cancer registries must be able to record systematically whether 'integrated' stage is truly integrated, or whether the origin of the component stage variables is simply unknown. Better international guidance is required for cancer registries on how to collect, prioritise and code data on stage at diagnosis. These points have been made in every paper that has emerged from the ICBP collaboration.

Given the quality of the available data, we find it difficult to imagine what else could have been done to improve the quality of the analyses on population-based cancer survival by stage at diagnosis, or the robustness of the interpretation of the findings.

C. Maringe, S. Walters, B. Rachet, J. Butler, P Finan, E Morris, A Gavin, MP Coleman

(1) Eden M, Rous B, Rashbass J. Misinterpretation of the origins and composition of staging data and its impact on colorectal cancer survival. Acta Oncol.
(2) Maringe C, Walters S, Rachet B, Butler J, et al. Stage at diagnosis and colorectal cancer survival in six high-income countries: A population-based study of patients diagnosed during 2000 – 2007. Acta Oncol 2013 June; 52(5):919-932
(3) Walters S, Maringe C, Butler J, Brierley JD, Rachet B, Coleman MP. Comparability of stage data in cancer registries in six countries: lessons from the International Cancer Benchmarking Partnership. Int J Cancer 2013 Feb 1;132(3):676-85.

Table 1: Origin of stage variable for patients included in survival analyses of the ICBP study: colon and rectal cancer patients diagnosed during 2000-7

	Valid Dukes' stage (%) [a]	*of which*	Dukes' stage value derived from T, N, M (%)	*of which*	Origin of 'T' code			Origin of 'N' code			Origin of 'M' code		
COLON					Pathological	Clinical	Unknown	Pathological	Clinical	Unknown	Pathological	Clinical	Unknown
Canada [b]	93.7		99.9		85.6	7.1	0.0	82.5	11.9	0.0	11.4	88.6	0.0
Denmark [b]	80.0		100.0		0.0	0.0	91.8	0.0	0.0	89.1	0.0	0.0	100.0
Sweden	96.6		30.1		0.0	0.0	92.8	0.0	0.0	91.1	0.0	0.0	100.0
UK	72.2		17.2		28.3	15.4	27.8	30.4	16.2	26.9	32.7	35.1	32.2
RECTUM													
Canada [b]	74.5		99.4		74.7	20.9	0.0	72.4	22.9	0.0	24.3	75.7	0.0
Denmark [b]	76.7		100.0		0.0	0.0	94.5	0.0	0.0	89.8	0.0	0.0	100.0
Norway [c]	70.0		0.0		-	-	-	-	-	-	-	-	-
Sweden	90.2		33.9		0.0	0.0	92.2	0.0	0.0	90.2	0.0	0.0	100.0
UK	69.4		18.8		31.1	13.0	29.3	32.2	14.5	28.7	31.1	34.1	34.8

[a] Final proportions of all patients for whom a valid stage category could be obtained from the raw data

[b] Patients diagnosed in 2004-7

[c] Final stage defined solely from Dukes' stage information: no component T, N and M codes available

Table 2: Comparisons of stage distribution (%) in the UK, by source

				Dukes' stage			
			Number	**A**	**B**	**C**	**D**
Eden et al.	**Colon and rectum combined**	"ICBP-like"	2,406	22	37	39	2
		Integrated	2,406	21	33	32	14
Maringe et al.[a]	**Colon**	**Final**	**102,555**	**9.4**	**38.6**	**35.1**	**16.9**
		T, N and M[b]	17,650	4.4	20.4	18.1	57.2
		Dukes[c]	84,905	10.5	42.4	38.7	8.5
	Rectum	**Final**	**46,769**	**20.7**	**28.6**	**33.8**	**16.9**
		T, N and M[b]	8,808	11.7	17.8	22.1	48.4
		Dukes[c]	37,961	22.8	31.1	36.5	9.6

[a] Observed stage distributions; imputation of missing stage values did not alter the distribution much

[b] Restriction to patients whose final stage was derived from the component T, N and M values

[c] Retsriction to patients whose final stage was derived from the raw Dukes' stage variable