Predictive model for risk of SLN involvement in EBC

Malcolm Hudson[a, b]

[a]NHMRC Clinical Trials Centre, Sydney University; [b]Dept. Statistics, Macquarie University

This version was compiled on December 11, 2018

The model was developed for Australian patients with early breast cancer (EBC) using data in the Sentinal Node Axillary Clearance (SNAC) randomized clinical trial, and validated in the Royal Adelaide Hospital (RAH).

Eligibility criteria in our modelling excluded women with ductal carcinoma in situ (DCIS), and those with outcome or tumour size not recorded.

1. Data sets

1.1. SNAC (n=982).

- SLNs were detected in 1024 of 1088 patients in the SNAC trial, of whom 982 were eligible for the present analysis. Exclusions were n=8 patients without SLN outcome, 34 patients with ductal carcinoma in situ (`tumtype` "DCIS") and 8 with missing information about tumour size (`ptumsize` NA).

- **snacVarnames**: *randnumb*, *metsize*, *positivy*, *ageyrs*, *metmm*, *ptumsize*, *tumourcm*, *lymphinv*, *septype*, *tumourtype*, *erpos*, *prpos*, *hgrade*, *MLTF*, *tumsite*, *metstage*, *metpos* and *macromets*

- Not considered in modelling: variables `septype` (2 levels, 113 positive), `method`(3 levels), `tumdet` (2 levels), `extinsitu` (2 levels, 159 positive) and `cadhis` (2 levels, only 29 positives). - Exclusions: SNAC model cohort is n=982 (of total n=1016) after excluding 1 case with unknown SLN status and all cases with tumour type DCIS.

1.2. RAH (n=541).

- **RAHvars**: *UR*, *age*, *presentation*, *palpability*, *size*, *site*, *type_tumour*, *multifocality*, *ER*, *PR*, *grade*, *LVI*, *mib_count*, *HER2*, *SLN_status*, *tumsite*, *tumtype*, *hgrade*, *lsize* and *LP*

- RAH variable `UR` is PID. - Maximum `size` is 50 (mm).

- Maximum number of values imputed in RAH is 15.
  - `mib_count` has 9 NAs, `tumsite` 12 NAs.

- some RAH variables (`HER2`, `mib_count`) not available for SNAC.
  - these can be used to test their ability to improve the SNAC derived model
  - an external validation of the SNAC model

2. Methods

2.1. Handling missing.

- missing SNAC data imputed as in Paper 1
- generally only a handful missing per variable, imputed by median or modal class
- an exception: in 60+ cases `prpos` missing, but a tight relationship exists with `erpos` and `hgrade`
- in this case we imputed missing `prpos` using this relationship

2.2. Modelling.

2.2.1. *Outcome.* `positivy` (SLN-status) is primary analysis.

- secondary `macromets`, similar fitting approach.

2.2.2. *Risk model.* Approach:

1. Develop the risk model in SNAC, using two approaches:
   (a) backward elimination approach
   (b) step up from Coombs model
   (c) Non-linearity was explored in either approach by introducing log-size in addition to size for logistic modelling and assessed graphically by examining relationship between logit(p) and tumour size in subgroups at otherwise homogeneous risk (e.g. data classified by lymph invasion status)
2. Internally validate the model developed (using bootstrap replicated datasets)
   (a) Harrell's internal bootstrap validation (ensuring model variables add additional discriminatory power) and
   (b) validation plots, and others
3. External validation in RAH
   (a) Before validating in RAH, option to permit simple (re)calibration to match rates of SLN-status in the RAH population.
   (b) We recommend recalibration for other large patient population in future application of our model.
   (c) Evaluate contributions of `HER2`, `palpability` and `mib_count` after adjusting for the new risk score. These variables are available only at RAH.
4. After validating in RAH, combine datasets and refit model for future use.

3. Results

3.1. Univariate associations. Factors were examined for univariate association with SLN-status. Table 1 shows univariate associations of risk variables with SLN-status outcome.

Dependent: positivy		0	1	p
tumsite	UOQ	344 (69.9)	148 (30.1)	0.044
	UIQ	138 (79.8)	35 (20.2)	
	Central	71 (65.7)	37 (34.3)	
	LIQ	62 (77.5)	18 (22.5)	
	LOQ	90 (69.8)	39 (30.2)	
ageyrs	Mean (SD)	57.9 (10.2)	56.2 (10.1)	0.016
lymphinv	N	610 (79.8)	154 (20.2)	<0.001
	Y	95 (43.6)	123 (56.4)	
hgrade	Grade 1	239 (77.1)	71 (22.9)	0.023
	Grade 2	303 (70.8)	125 (29.2)	
	Grade 3	163 (66.8)	81 (33.2)	
tumpal	0	341 (82.0)	75 (18.0)	<0.001
	1	364 (64.3)	202 (35.7)	
prpos	N	197 (73.2)	72 (26.8)	0.537
	Y	508 (71.2)	205 (28.8)	
erpos	N	128 (72.7)	48 (27.3)	0.761
	Y	577 (71.6)	229 (28.4)	
MLTF	N	663 (72.2)	255 (27.8)	0.257
	Y	42 (65.6)	22 (34.4)	
tumtype	Ductal	580 (71.8)	228 (28.2)	0.749
	Lobular	66 (69.5)	29 (30.5)	
	other	59 (74.7)	20 (25.3)	
ptumsize	Mean (SD)	14.4 (7.3)	20.8 (11.5)	<0.001
ltumsize	Mean (SD)	2.9 (0.6)	3.4 (0.6)	<0.001
LP_C	Mean (SD)	-1.1 (0.6)	-0.4 (1.4)	<0.001

Table 1. Univariate associations

Dependent: positivy		0	1	OR (univariable)	OR (multivariable)
tumsite	UOQ	344 (48.8)	148 (53.4)	-	-
	UIQ	138 (19.6)	35 (12.6)	0.59 (0.38-0.89, p=0.013)	0.57 (0.36-0.90, p=0.018)
	Central	71 (10.1)	37 (13.4)	1.21 (0.77-1.87, p=0.395)	1.20 (0.74-1.94, p=0.451)
	LIQ	62 (8.8)	18 (6.5)	0.67 (0.38-1.16, p=0.168)	0.69 (0.37-1.25, p=0.234)
	LOQ	90 (12.8)	39 (14.1)	1.01 (0.66-1.53, p=0.973)	0.95 (0.59-1.49, p=0.817)
ageyrs	Mean (SD)	57.9 (10.2)	56.2 (10.1)	0.98 (0.97-1.00, p=0.022)	-
lymphinv	N	610 (86.5)	154 (55.6)	-	-
	Y	95 (13.5)	123 (44.4)	5.13 (3.73-7.08, p<0.001)	3.92 (2.76-5.58, p<0.001)
hgrade	Grade 1	239 (33.9)	71 (25.6)	-	-
	Grade 2	303 (43.0)	125 (45.1)	1.39 (0.99-1.95, p=0.056)	0.79 (0.54-1.16, p=0.236)
	Grade 3	163 (23.1)	81 (29.2)	1.67 (1.15-2.44, p=0.007)	0.68 (0.44-1.06, p=0.088)
tumpal	0	341 (48.4)	75 (27.1)	-	-
	1	364 (51.6)	202 (72.9)	2.52 (1.87-3.43, p<0.001)	1.71 (1.21-2.43, p=0.003)
prpos	N	197 (27.9)	72 (26.0)	-	-
	Y	508 (72.1)	205 (74.0)	1.10 (0.81-1.52, p=0.537)	-
erpos	N	128 (18.2)	48 (17.3)	-	-
	Y	577 (81.8)	229 (82.7)	1.06 (0.74-1.54, p=0.761)	-
MLTF	N	663 (94.0)	255 (92.1)	-	-
	Y	42 (6.0)	22 (7.9)	1.36 (0.78-2.30, p=0.258)	-
tumtype	Ductal	580 (82.3)	228 (82.3)	-	-
	Lobular	66 (9.4)	29 (10.5)	1.12 (0.69-1.76, p=0.637)	-
	other	59 (8.4)	20 (7.2)	0.86 (0.50-1.44, p=0.584)	-
ptumsize	Mean (SD)	14.4 (7.3)	20.8 (11.5)	1.08 (1.06-1.10, p<0.001)	1.06 (1.04-1.08, p<0.001)
ltumsize	Mean (SD)	2.9 (0.6)	3.4 (0.6)	3.35 (2.61-4.36, p<0.001)	-
LP_C	Mean (SD)	-1.1 (0.6)	-0.4 (1.4)	2.68 (2.13-3.42, p<0.001)	-

Table 2. Univariate associations and stepwise elimination.

3.2. Multivariable risk model development. Table 2 shows (under "multivariable") the stepwise reduced multivariate model using a liberal inclusion P-value criterion (P-value to retain < 0.20) from a risk variable stock including tumour size and `hgrade` but *not* including the Coombs risk score.

The null deviance (the model lack of fit statistic) in SNAC provided by the Coombs model is 1074.9.

The lrm() fitting formula (positivy ~ tumsite + ageyrs + lymphinv + hgrade + tumpal + prpos + erpos + MLTF + tumtype + ptumsize + ltumsize + LP_C + offset(LP_C)) reduces this lack of fit to deviance 985.9.

The alternative model – employing approach 1(b) – developed by forward logistic model from the Coombs model (as offset) demonstrates that Coombs adjustments for `hgrade` are not prognostic in SNAC.

First, logistic forward modelling *adding factors to the Coombs predicted probability* adds lymph-invasion (`lymphinv`) status and adjusts `hgrade` coefficients significantly, while introducing other factors, particularly tumour palpability `tumpal` associated with `hgrade`:
see Table 3.

	Estimate	Std. Error	z value	Pr(>|z|)
lymphinvN	-1.58	1.02	-1.55	0.12
lymphinvY	-0.23	1.05	-0.22	0.83
hgradeGrade 2	-0.53	0.20	-2.70	0.01
hgradeGrade 3	-0.86	0.23	-3.79	0.00
tumpal	0.47	0.18	2.53	0.01
tumsiteUIQ	-0.60	0.24	-2.52	0.01
tumsiteCentral	0.17	0.25	0.68	0.49
tumsiteLIQ	-0.41	0.31	-1.33	0.18
tumsiteLOQ	-0.09	0.24	-0.36	0.72
ptumsize	-0.07	0.03	-2.16	0.03
ltumsize	0.83	0.51	1.64	0.10
MLTFY	0.48	0.31	1.56	0.12

Table 3. Summary of fwd.glm

Second, the corresponding model positivy ~ lymphinv + hgrade + tumpal + tumsite + ptumsize + ltumsize + MLTF which includes *all* Coombs risk factors but *no* Coombs probability offset is simplified by removing `hgrade` without statistically significant deterioration in fit ($\chi^2 = 14.4$ on 8 degrees of freedom (P=0.07) after dropping `hgrade`, `MLTF`, `tumsite` and *untransformed* tumour size).

The ANOVA of Table 4 also demonstrates the lack of contribution.

	Chi-Square	d.f.	P
lymphinv	56.09	1	0.000
hgrade	4.01	2	0.135
tumpal	6.23	1	0.013
tumsite	8.54	4	0.074
ptumsize	0.16	1	0.692
ltumsize	3.11	1	0.078
MLTF	2.04	1	0.153
TOTAL	142.34	11	0.000

Table 4. ANOVA of fmla1.lrm

Logarithm of tumour size empirically (`checkFunctionalForm`, not included here) provides a linear relationship with logit(p) and is most strongly associated in models.

In short, we have found the Coombs model does not fit SNAC data adequately, and is improved on by *removing* `hgrade` from prediction while adopting a logistic model depending only on lymph invasion status, `lymphinv`, tumour palpability, `tumpal`, and log-transformed tumour size, `ltumsize`.

The resulting multivariable model developed in SNAC is that of Table 5

Dependent: positivy		0	1	OR (univariable)	OR (multivariable)
lymphinv	N	610 (86.5)	154 (55.6)	-	-
	Y	95 (13.5)	123 (44.4)	5.13 (3.73-7.08, p<0.001)	3.57 (2.55-5.02, p<0.001)
hgrade	Grade 1	239 (33.9)	71 (25.6)	-	-
	Grade 2	303 (43.0)	125 (45.1)	1.39 (0.99-1.95, p=0.056)	-
	Grade 3	163 (23.1)	81 (29.2)	1.67 (1.15-2.44, p=0.007)	-
tumpal	0	341 (48.4)	75 (27.1)	-	-
	1	364 (51.6)	202 (72.9)	2.52 (1.87-3.43, p<0.001)	1.43 (1.01-2.02, p=0.046)
ptumsize	Mean (SD)	14.4 (7.3)	20.8 (11.5)	1.08 (1.06-1.10, p<0.001)	-
ltumsize	Mean (SD)	2.9 (0.6)	3.4 (0.6)	3.35 (2.61-4.36, p<0.001)	2.38 (1.79-3.18, p<0.001)

Table 5. SNAC multivariable model

3.3. Internal validation in SNAC. The expected number of SLN_status positives is estimated as 277 of $n = 982$ in SNAC. The observed number is 277 in $n = 982$.

Stepwise reduction (using R functions `rms::fastbw()`) in 100 bootstrap data resamples (generated in `rms::validate()`) provides

```
#
#       Backwards Step-down - Original Model
#
#   Deleted  Chi-Sq d.f. P      Residual d.f. P      AIC
#   ptumsize 0.16   1    0.6924  0.16    1    0.6924 -1.84
#   MLTF     2.06   1    0.1512  2.22    2    0.3301 -1.78
#   hgrade   3.74   2    0.1544  5.95    4    0.2027 -2.05
#   tumsite  7.96   4    0.0930 13.91    8    0.0840 -2.09
#
#  Approximate Estimates after Deleting Factors
#
#                Coef    S.E. Wald Z         P
#  Intercept  -4.1369 0.4389 -9.425 0.000e+00
#  lymphinv=Y  1.2491 0.1742  7.171 7.425e-13
#  tumpal      0.3546 0.1790  1.981 4.756e-02
#  ltumsize    0.8477 0.1456  5.821 5.852e-09
#
#  Factors in Final Model
#
#  [1] lymphinv tumpal   ltumsize
```

```
#  lymphinv   hgrade   tumpal  tumsite ptumsize ltumsize     MLTF
#       100       24       54       55       20       80       21
```

`lympinv` was selected in all 100 models, either `ltumsize` (over 80 models) or `ptumsize` (under 20 models) in all 100 cases, with `hgrade` selected also in some models.

3.4. External validation in RAH.

3.4.1. Calibration. The association of our newly developed risk score with SLN status was C=0.70,maintaining a similar level of discriminatory ability as it had in SNAC.

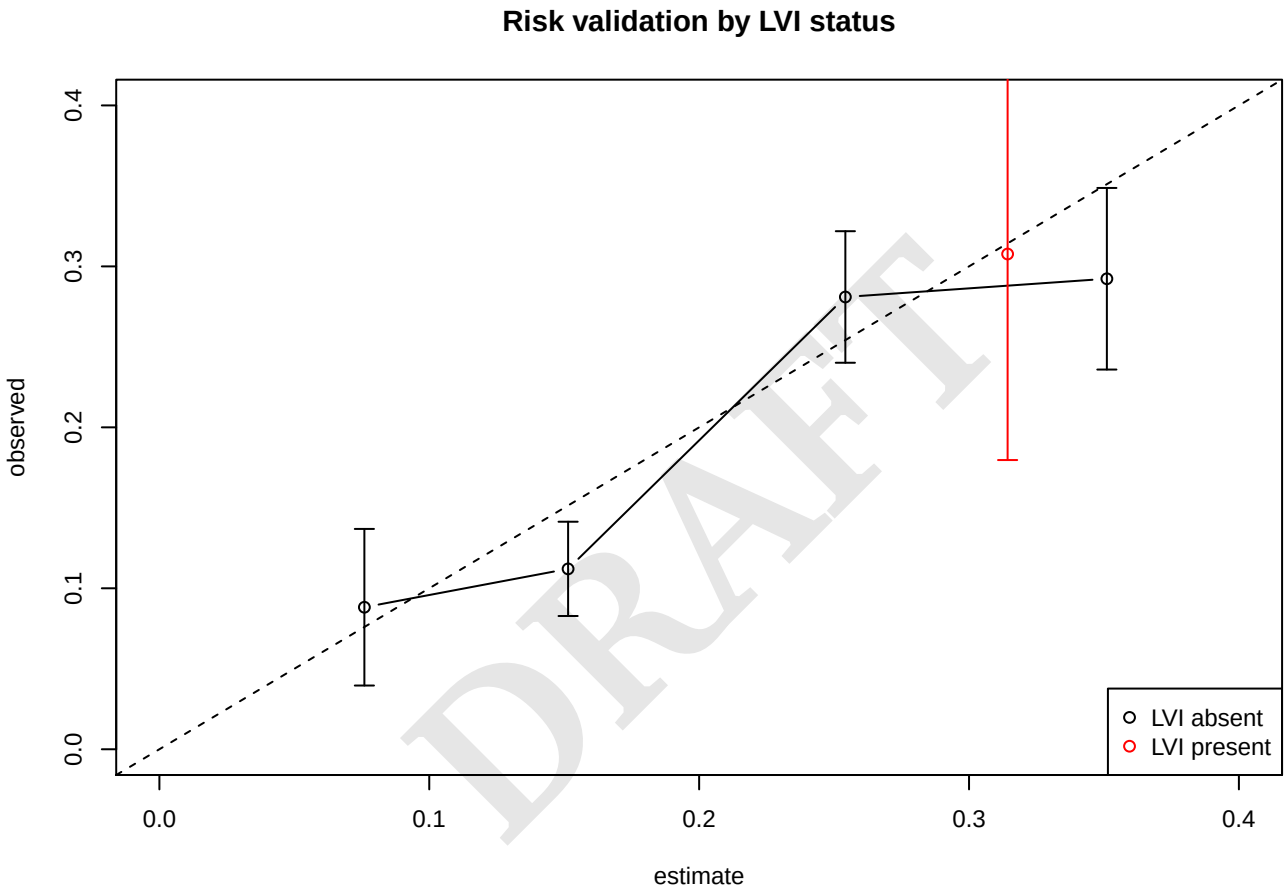
While the C-index of the association is strong the SNAC developed model failed to calibrate at RAH over the whole range of risks encountered. This is indicate in the tests of intercept (z=-2.79, P=0.005) and slope (z=-2.11, P=0.02) in the fit between the risk score LP and outcome.

However, the failure of calibration is predominantly in predictions for *high* risk women.

The number of women with SLN_status positive at RAH was 172 of $n = 541$ (31.79 %). The risk factor LVI status was undetermined for one woman and palpability undetermined for two women.

The SNAC risk score, uncalibrated, estimated 191 of 538 women would be SLN_status positive, a slight over-estimate. Of n=373 women with LVI absent, the numbers SLN status positive, observed and predicted, were 83 and 92. Only 13 of 164 women with LVI present had probability predictions below 0.4, so presence of LVI would indicate SLN biopsy in nearly all such women.

Calibration appeared quite good for probabilities below 0.4, the region important for the decision tool. No adjustment for calibration in the large or shrinkage was therefore required.



3.4.2. Benefits of predictors unique to RAH. HER status, mib count conferred no improvement on the SNAC developed model.