

Keywords: analytic methods; chemoresponse assay; chemosensitivity assay; predictive marker; personalised medicine

Evaluation of chemoresponse assays as predictive markers

E L Korn^{*,1} and B Freidlin¹

¹Biometric Research Branch, MSC 9735, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD 20892, USA

Background: A chemoresponse assay that can be used to predict which patients will respond to which drugs would be useful in directing treatment. Two new analytic methods to assess the predictive ability of a chemoresponse assay have been proposed by Tian *et al.*

Methods: Three examples in which a hypothetical assay has no predictive ability are considered to evaluate the properties of the proposed analytic methods.

Results: For these specific examples, the proposed methods incorrectly suggest that the assay is predictive.

Conclusions: The examples presented here demonstrate that it can be challenging to evaluate the predictive value of a chemoresponse assay.

The statistical evaluation of non-randomised chemoresponse assay data is nontrivial because the complex interrelationship between differing efficacies of the therapies and differing underlying prognoses of the patients can lead to an assay appearing to have predictive ability when it does not (Wieand, 2005). For example, an analysis of retrospective assay results that compares the clinical outcomes of patients who received an agent for whom their assay predicts sensitivity *vs* the outcomes of patients whose assay result predicted resistance will show a difference in average outcomes if the assay is solely prognostic, that is, the assay-sensitive results are associated with good outcomes regardless of what treatment the patients receives. Tian *et al* (2014) proposed two new analytic methods applied to assay-response data intended to assess the predictive ability of an assay (biomarker), that is, whether the assay can predict response to a particular therapy *vs* other therapies. We evaluated whether the proposed methods are reliable.

MATERIALS AND METHODS

For the ‘match/mismatch analysis’ of Tian *et al* (2014), the assay–outcome association is calculated for the observed data (match analysis) and for a permuted version of the data in which the assay result (sensitive or resistant) for each patient is randomly selected

(from all the assay results for that patient) when estimating the assay–outcome association (mismatch analysis). If the match association is stronger than the mismatch association, Tian *et al* (2014) suggest that the assay is predictive. For the ‘cross-drug response’ analysis, the patients are divided into four groups based on the assay’s predictions of sensitivity and the actual therapy received: SA (assay sensitive to all therapies), SP (assay sensitive to some treatments and treated with a sensitive therapy), RA (assay resistant to all therapies), and RP (assay resistant to some therapies and treated with a resistant therapy). If the average outcomes for SA and SP are better than the average outcomes for RA and RP, and the outcomes for SA and SP are similar, and the outcomes for RA and RP are similar, then Tian *et al* (2014) suggest the assay is predictive.

Based on our understanding of how these analytic methods are applied, we assess whether these methods are reliable via three hypothetical examples. The use of hypothetical examples is a useful way to assess whether an analytic method works because, unlike using observed data, one knows the true state of nature. The examples use response rates (RRs) rather than progression-free survival as the outcome (as was done by Tian *et al*, 2014) because this makes all the calculations transparent and easily verifiable with a hand calculator. The examples are presented to critically assess the analytic methods, and are not intended to reflect on the particular chemoresponse assay evaluated by Tian *et al* (2014; Rutherford *et al*, 2013).

*Correspondence: Dr EL Korn; E-mail: korne@ctep.nci.nih.gov

Received 10 October 2014; revised 4 December 2014; accepted 18 December 2014; published online 13 January 2015

© 2015 Cancer Research UK. All rights reserved 0007–0920/15

RESULTS

Consider the example described in Table 1 for two treatments. One-half the population is treated with A and one-half with B. The important point to notice about the RRs in Table 1 is that, for each row of the table, they are the same for treatment A and treatment B. Therefore, the assay results are not predictive. That is, the treatment effect (difference in RRs) is the same regardless of assay results. In addition, using the assay to direct treatment could not improve overall RRs; the overall observed RR is 38%, which is the same as would be obtained if the assay directed treatment. Now consider the match/mismatch analysis of Tian *et al* (2014). The RRs for those patients who received a treatment for which the assay suggested sensitivity *vs* resistance to that treatment (the ‘match analysis’) are (see the Supplementary Appendix):

- Observed assay result of sensitive-to-treatment RR = 45%
- Observed assay result of resistant-to-treatment RR = 10%
- Difference = 35%

On the other hand, the analysis with a randomly selected assay result (the ‘mismatch analysis’) yields:

- Random assay result of sensitive-to-treatment RR = 42.67%
- Random assay result of resistant-to-treatment RR = 24.00%
- Difference = 18.7%

As the assay–outcome association is smaller for the mismatch analysis (18.7%) than for the match analysis (35%), Tian *et al* (2014) would incorrectly suggest that this assay is predictive.

For the cross-drug response analysis,

- SA (assay sensitive to all therapies): RR = 45%
- SP (assay sensitive to some treatments and treated with a sensitive therapy): RR = 45%
- RA (assay resistant to all therapies): RR = 10%
- RP (assay resistant to some therapies and treated with a resistant therapy): RR = 10%

Although the data in Table 1 perfectly satisfies the criteria of Tian *et al* (2014) for being predictive using this analysis, as noted above the assay has zero predictive ability.

In Table 1, the treatment assignment is not random, with patients with better prognoses receiving A (as can be seen by the higher RRs, these patients would have had regardless of what treatment they receive). If one is willing to assume that the treatments the patients received were given essentially at random, then, in theory, it may be easier to evaluate an assay. However, the analyses suggested by Tian *et al* (2014) can lead to confusing results even in this situation. For example, to see that the match/mismatch analysis can be misleading, consider the hypothetical data in Table 2, for which the treatment assignment is randomly chosen (i.e., the distribution of assay results and RRs are identical for patients who received A *vs* B). For the ‘match analysis’:

- Observed assay result of sensitive-to-treatment RR = 52.73%
- Observed assay result of resistant-to-treatment RR = 26.67%
- Difference = 26.06%

and for the ‘mismatch analysis’:

- Random assay result of sensitive-to-treatment RR = 47.27%
- Random assay result of resistant-to-treatment RR = 33.33%
- Difference = 13.94%

As the assay–outcome association is smaller for the mismatch analysis (13.94%) than the match analysis (26.06%), Tian *et al* (2014) would incorrectly suggest that the assay is predictive. Note that if B was the standard treatment in this situation where A is uniformly better than B, use of the assay could improve the overall RR because more patients would be treated with A. However, this does not represent a predictive value of the assay; in fact, even a better overall RR is achieved by everyone being treated with A.

To see how even with random treatment assignment the cross-drug response analysis can yield the wrong conclusion, consider the example in Table 3. Similar to Table 2, this example has the treatments (in this case, three of them) being assigned to patients at random (as can be seen by the same numbers in the three horizontal panels of the table). Note that the assay has no predictive ability: for each assay category, the RR for A is 15% higher than for B, which is 15% higher than for C. For the cross-drug response analysis

- SA (assay sensitive to all therapies): RR = 50%

Table 1. Hypothetical example 1: Response rates to two treatments (A and B) stratified by which treatment patients would typically receive in the population and assay results (proportions in parenthesis are the proportions of patients in the population in each category)

Subpopulation treated with:	Assay results			Response rate if patients treated with:	
	A	B		A (%)	B (%)
A (0.5)	Sensitive	Sensitive	(0.2)	75	75
	Sensitive	Resistant	(0.2)	45	45
	Resistant	Resistant	(0.1)	10	10
B (0.5)	Sensitive	Sensitive	(0.4)	30	30
	Sensitive	Resistant	(0.1)	10	10

Percentages in bold are response rates that would be observed. Assay result combinations not appearing in the table do not occur for this hypothetical example.

Table 2. Hypothetical example 2: Response rates to two treatments (A and B) stratified by which treatment patients would typically receive in the population and assay results (proportions in parenthesis are the proportions of patients in the population in each category)

Subpopulation treated with:	Assay results			Response rate if patients treated with:	
	A	B		A (%)	B (%)
A (0.5)	Sensitive	Sensitive	(0.20)	70	30
	Sensitive	Resistant	(0.15)	60	20
	Resistant	Resistant	(0.15)	50	10
B (0.5)	Sensitive	Sensitive	(0.20)	70	30
	Sensitive	Resistant	(0.15)	60	20
	Resistant	Resistant	(0.15)	50	10

Percentages in bold are response rates that would be observed. Assay result combinations not appearing in the table do not occur for this hypothetical example.

Table 3. Hypothetical example 3: Response rates to three treatments (A, B, and C) stratified by which treatment patients would typically receive in the population and assay results (proportions in parenthesis are the proportions of patients in the population in each category)

Subpopulation treated with:	Assay results				Response rate if patients treated with:		
	A	B	C		A (%)	B (%)	C (%)
A (1/3)	Sensitive	Sensitive	Sensitive	(1/12)	65	50	35
	Sensitive	Sensitive	Resistant	(1/12)	60	45	30
	Sensitive	Resistant	Resistant	(1/12)	45	30	15
	Resistant	Resistant	Resistant	(1/12)	40	25	10
B (1/3)	Sensitive	Sensitive	Sensitive	(1/12)	65	50	35
	Sensitive	Sensitive	Resistant	(1/12)	60	45	30
	Sensitive	Resistant	Resistant	(1/12)	45	30	15
	Resistant	Resistant	Resistant	(1/12)	40	25	10
C (1/3)	Sensitive	Sensitive	Sensitive	(1/12)	65	50	35
	Sensitive	Sensitive	Resistant	(1/12)	60	45	30
	Sensitive	Resistant	Resistant	(1/12)	45	30	15
	Resistant	Resistant	Resistant	(1/12)	40	25	10

Percentages in bold are response rates that would be observed. Assay result combinations not appearing in the table do not occur for this hypothetical example.

SP (assay sensitive to some treatments and treated with a sensitive therapy): RR = 50%

RA (assay resistant to all therapies): RR = 25%

RP (assay resistant to some therapies and treated with a resistant therapy): RR = 25%

This analysis incorrectly suggests that the assay is predictive (see the discussion of Table 2).

CONCLUSIONS

Given the limitations of the two approaches suggested by Tian *et al* (2014), what do we recommend for analysis of observational chemoresponse assay studies? If it is not reasonable to assume that the treatments the patients received were assigned essentially randomly (at least approximately), then the likelihood of being able to draw reliable conclusions about the predictive ability of the assay appears remote. If one is willing to make the randomness assumption, then one could evaluate the predictive value of the assay for each pair of therapies in a straightforward manner by examining the outcomes stratified by the assay results for that pair of therapies. (One can then get an estimate of the overall predictive utility of the assay by integrating the conclusions over all pairs of therapies, e.g., as Korn *et al* (1985) did for a prognostic assay). For example, for Table 2, the treatment effects (increases in RR) that would be observed in the sensitive–sensitive, sensitive–resistant, and resistant–resistant populations are all 40%, correctly suggesting that the assay is not predictive. Conclusion from such an evaluation would need to be tentative, because of the uncertainty of the randomness assumption. A biomarker-strategy design in which patients are randomly assigned to assay-directed therapy *vs* standard therapy is definitive, but is problematic (Grendys *et al*, 2014). When a large proportion of patients are expected to receive

a limited number of treatments, a definitive predictive evaluation can be performed using a biomarker-stratified randomised trial. In this design, patients are randomly assigned among these treatments, and the different treatments are evaluated within each biomarker-assay category (Freidlin *et al*, 2010).

REFERENCES

- Freidlin B, McShane LM, Korn EL (2010) Randomized clinical trials with biomarkers: design issues. *J Natl Cancer Inst* **102**: 152–160.
- Grendys Jr EC, Fiorica JV, Orr Jr JW, Holloway R, Wang D, Tian C, Chan JK, Herzog TJ (2014) Overview of a chemoresponse assay in ovarian cancer. *Clin Transl Oncol* **16**: 761–769.
- Korn EL, Sondak VK, Bertelsen CA, Kern DH (1985) Analysis of the clinical utility of a predictive chemosensitivity assay. *Stat Med* **4**: 527–534.
- Rutherford T, Orr Jr J, Grendys Jr E, Edwards R, Krivak TC, Holloway R, Moore RG, Puls L, Tillmanns T, Schink JC, Brower SL, Tian C, Herzog TJ (2013) A prospective study evaluating the clinical relevance of a chemoresponse assay for treatment of patients with persistent or recurrent ovarian cancer. *Gynecol Oncol* **131**: 362–367.
- Tian C, Sargent DJ, Krivak TC, Powell MA, Gabrin MJ, Brower SL, Coleman RL (2014) Evaluation of a chemoresponse assay as a predictive marker in the treatment of recurrent ovarian cancer: further analysis of a prospective study. *Br J Cancer* **111**: 843–850.
- Wieand HS (2005) Chemotherapy sensitivity and response assays: are the ASCO guidelines for clinical trial design too restrictive? *J Clin Oncol* **23**: 3643–3644.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 4.0 Unported License.

Supplementary Information accompanies this paper on British Journal of Cancer website (<http://www.nature.com/bjc>)