Author’s response to reviews

Title: Validation of a registry-derived risk algorithm based on treatment protocol as a proxy for disease risk in childhood acute lymphoblastic leukemia

Authors:

Sumit Gupta (sumit.gupta@sickkids.ca)
Jason D Pole (jpole@pogo.ca)
Astrid Guttmann (astrid.guttmann@ices.on.ca)
Lillian Sung (lillian.sung@sickkids.ca)

Version: 2 Date: 18 April 2013

Author’s response to reviews:

Sumit Gupta, MD
Division of Haematology/Oncology
The Hospital for Sick Children
555 University Avenue
Toronto, Ontario
M5G 1X8

April 15th, 2013

Adrian Aldcroft
Executive Editor
BMC Medical Research Methodology

Dear Editors,

We would like to thank you and the reviewers for your thoughtful consideration of our manuscript, which we believe has benefited greatly as a result. Please find below an itemized list of the reviewer comments and our responses. The respective changes to the manuscript have also been highlighted in yellow in the document for ease of viewing.

We look forward to further correspondence with your office.

Sincerely yours,

Dr. Sumit Gupta

REVIEWER 1:

This was a well written manuscript that proposes important options for utilizing data from a cancer registry. Studies such as these are necessary as they can
extend the impact of these data sources to important clinical questions that may not be answered from smaller trial datasets.

Thank you.

Minor essential revisions:

1. One could argue that this was not a validation study but rather a derivation of a registry based risk algorithm—as they compared their registry algorithm to 4 different definitions for high risk based on age/lab/biologic data. Seems that to truly be a validation study they would have chosen a single “gold standard” a priori and confirmed the validity of the registry against that gold standard. The approach taken seemed to explore the correlation of the registry approach to various biology-based algorithms, which is fine to do just may not be a true validation study.

The reviewer is correct in that traditional validation studies in health services research (HSR) define a gold standard. In our study, we attempted to determine whether registry-derived treatment protocol name accurately reflected biology-based disease risk. As there are multiple risk classification systems in acute lymphoblastic leukemia of increasing complexity, we compared our registry derived risk algorithm to each instead of choosing a single gold standard. Nonetheless, we believe that approach could be considered a form of criterion validity, since the criterion has indeed changed over time.

2. Not sure why MRD was considered in the final biology-based algorithm. I understand that this is an important component for risk stratification during the course of treatment but it is not something that is known at baseline when the initial chemo choice is made so wonder if it is appropriate to include. Furthermore, this element was missing for half of the patients making it relatively uninformative. Did the authors reflex patients with missing MRD to negative? If so this would not be an appropriate approach. Would argue removing this 4th algorithm from the analysis—especially since the addition of this data element made no apparent difference. If MRD is thought to be necessary to the designation of high risk status in the biology algorithm then the authors would need to find a cohort where this data element is more complete.

The reviewer makes an excellent point. While MRD is not available at diagnosis, it is nonetheless used for risk stratification in virtually all current studies examining ALL outcomes. Thus our question was whether our protocol based algorithm could validly approximate disease risk stratifications systems which included MRD. Out of the same concern expressed by the reviewer, we did not restrict ourselves to this most complex model, but also examined disease risk classifications that did not include MRD. In cases where MRD was not available, it was treated as non-informative. A sentence to this effect has been added to the methods as follows on page 7

“Where a prognosticator was not available for a particular patient, the algorithm treated it as non-informative.”
3. Would like to have had a little more info on how patients were linked between the two data sources. It appears that the linking process was quite successful but may be of interest to the reader to know how this was done.

Registry personnel situated at each treatment center assign patients unique identifiers at diagnosis. These numeric identifiers are retained by the treating centers and are therefore available via chart review. These identifiers were used for linkage. To make this process clearer, we have changed the relevant section in the methods to read the following:

“POGONIS is a population-based registry that prospectively captures all cases of pediatric cancer diagnosed and treated at one of the five tertiary pediatric oncology centers in Ontario. POGONIS personnel assign each patient a unique numeric identifier, which is retained both by the treating centers and POGONIS. This number was therefore used to link study patients to POGONIS.”

4. The authors state that the agreement was “almost perfect”. Not sure I would use these words. Hard to quantify what “almost perfect” means—I realize that this comes from a prior cited publication but seems to me that these designations given in the 1977 article were that author’s preference.

The reviewer is absolutely correct in that “almost perfect” was the term first suggested by Landis and Koch to describe kappas greater than 0.80, but was not justified and is arguably arbitrary. Other (equally arbitrary) classifications have been also been suggested. In all cases, 0.75 or 0.80 have been the lower border of the highest category. Therefore, we have changed “almost perfect” to “excellent” and added a second reference for a second classification system.

5. Not sure that the kappa statistic was the best analytic approach. Certainly it tells you how well standard designations and high risk designations from the chart review match up with the biology designations but it does not well represent the mismatches—was the registry algorithm misclassifying standard risk or high risk patients? The authors do give a list of each of the scenarios that did not match up in the text but this is difficult to follow and the reader can get lost in the text. Why not establish false positive/false negative frequencies and other operating characteristics such as sens, spec, PPV, NPV? These operating characteristics could easily be presented in a table and thus variations across each of the 4 algorithms could more easily be evaluated. This approach is still simple and would be much more informative then a single kappa statistic which is limited.

The kappa statistic does in fact take into consideration mismatches but further, is better than percent agreement as it takes into consideration agreement by chance. However, we agree with the reviewer’s point related to including operating characteristics. We have calculated the parameters suggested and added them to Table 4 as requested.

6. The authors appropriately discuss the limitations of their approach—the most important limitation is the lack of generalizability of their approach to other
registries or administrative data that may not contain similar data elements to this registry (i.e. this registry benefited from having specific protocol designations that other registries may or may not have). In the end this paper supports the use of this risk designation algorithm for only this registry and only during the time of study (as the biologic designation of risk seems to be changing over time)—however, it is important to do this process in order to use the registry for additional analyses requiring risk stratification. The authors should highlight that their specific approach likely is not applicable to other registries but that other researches need to carry out similar approaches to be able to designate risk status from a registry or administrative database.

We agree with the reviewers point. In the discussion, we comment on the limitation in our ability to generalize our findings “to different malignancies and treatment protocols”. We also state that “in addition, some population-based registries may not collect treatment protocol names” and that other ways of defining treatment intensity may be useful but require validation. To address the reviewer’s specific points, we have added the following sentence to the end of this paragraph:

“Even where other registries or administrative databases do collect protocol name, validation analyses similar to that carried out in this study are likely to be necessary before they can be used to carry out analyses requiring risk stratification.”

7. Minor edit: Second sentence of intro “…reported in from clinical trials...” needs to be fixed.

This has been corrected in the manuscript.

REVIEWER 2:

Discretionary Revisions

1. The findings presented here are sound for the specific scenario analyzed, although their generalizability is unclear, since treatment protocol is not always a surrogate marker for risk group as it was here, either in pediatric ALL or in other malignancies. Treatment stratum is often a better indicator of risk status than treatment protocol, since patients of varying risk groups are often treated on different strata within a single protocol. It would be informative if the authors could comment on whether or not treatment stratum is generally available in the registry. If so, inclusion of this information in the registry-based algorithm would be optimal. If not, the utility of a registry-based algorithm based only on protocol name would be limited whenever a treatment protocol contained strata for multiple risk groups.

Please see the response to Comment 6 of Reviewer 1 above. To address this specific point, we have added the following sentence to the discussion of generalizability:

“Some cooperative groups treat ALL with protocols whose names do not differ by
risk strata, or where multiple treatment strata are contained within a single protocol name."

2. It is not uncommon for risk group to be reassigned based on information obtained at a late time point (over 4 weeks following start of treatment), eg based on end-induction MRD (in ALL and AML), or surgical outcome (in many solid tumors). This did not cause problems in the current study because nearly half the patients were treated on older protocols that did not use MRD for risk assignment, and the remainder were treated on protocols that used a higher MRD cutoff of 0.1% than is presently used, which resulted in relatively few patients being reassigned from SR to HR based on MRD positivity. However, in the present era, a substantial number of SR patients are reassigned to HR based on MRD exceeding the 0.01% cutoff, so agreement between post-induction risk status and initial treatment protocol assignment would be poorer. Therefore for ALL patients treated on modern regimens, a use of registry-based algorithm would more accurately reflect risk status if it were based on the post-induction treatment protocol assignment rather than the protocol assigned during the first four weeks of therapy.

Please see response to Comment 2 of Reviewer 1 above.

Minor Essential Revisions

1. Third paragraph of discussion - new cytogenetic abnormalities – should add reference(s) re another recently identified lesion, iAMP21, in addition to ref 21.

   This reference has been added.

2. Table 1 – hypodiploidy – legend says >45 instead of <45 chromosomes

   This has been corrected in the table.