THIOPURINE S-METHYLTRANSFERASE TESTING FOR
AVERTING DRUG TOXICITY IN PATIENTS RECEIVING
THIOPURINES: A SYSTEMATIC REVIEW AND QUALITY
APPRAISAL

Report No. 2015-02
Date: July 29, 2015

REPORT HIGHLIGHTS

The Report Highlights consists of a summary of the full report with
the same name and should be evaluated in conjunction with the full
report and its appendices. Full documents are available for download
at:

http://lab.research.sickkids.ca/task/reports-theses/
Authors
Lilla M. Roy, RN, BScN, MSc, Child Health Evaluative Sciences, The Hospital for Sick Children
Wendy J. Ungar, MSc, PhD, Child Health Evaluative Sciences, The Hospital for Sick Children
Richard M. Zur, PhD, Child Health Evaluative Sciences, The Hospital for Sick Children

Co-Investigators
Joseph Beyene, MSc, PhD, Department of Clinical Epidemiology & Biostatistics, McMaster University
Chris Carew, MBA, Centre for Genetic Medicine, The Hospital for Sick Children
Shinya Ito, MD, FRCPC, Clinical Pharmacology and Toxicology, The Hospital for Sick Children,
Pharmacology & Pharmacy, Department of Paediatrics, University of Toronto
Elizabeth Uleryk, MLS, The Hospital for Sick Children Library, Toronto
James Whitlock, MD, Haematology/Oncology, The Hospital for Sick Children, Paediatrics, University of Toronto

Acknowledgements
This research was supported by a Canadian Institutes of Health Research Knowledge Synthesis Grant, grant #305352.

We thank the following individuals who provided assistance with study translation: Fernando Ortiz, Ella Hyatt, Zhaoxin Dong, Roland Arnold, Takashi Hirama, Cindy Zhang, Jessie Guo, Simon Tian, Brian Kim, Jennifer Park and Vera Nenadovic. We would also like to thank Tara Paton and Kassa Beimnet for their technical expertise regarding genotype and phenotype testing, and Cathy Pajunen for her library services and referencing technical support.

The views expressed in the material are the views of the authors and do not necessarily reflect those of The Hospital for Sick Children or the province of Ontario.

Conflicts of interest
The authors have no conflicts of interest to disclose.
Introduction

Thiopurine S-methyltransferase (TPMT) is an enzyme that metabolizes thiopurine drugs. Thiopurine drugs are commonly used in maintenance treatment for childhood leukemias, as well as, less commonly, for inflammatory bowel disease (IBD), transplant recipients, and dermatological conditions. The absence or a deficiency of TPMT can significantly increase the risk of adverse drug event (ADE) in persons receiving thiopurine therapy as they are unable to metabolize the drug.

Rationale

Phenotype testing for TPMT activity has traditionally been the test of choice to assess TPMT deficiency, however, the recent increase in capacity and use of genotype testing to assist in disease treatment has evoked questions about which method of testing is optimal.

Objectives

The objectives were to systematically review the evidence on the performance characteristics of phenotype and genotype testing for TPMT deficiency,

Key Messages

• Along with traditional phenotype testing, there is a growing use of personalized medicine applications such as pharmacogenomics in clinical diagnostics and clinical decision-making for selection of drug treatment and dose.
• Thiopurine S-methyltransferase (TPMT) testing can use either phenotype or genotype methods. Both tests have limitations, and sensitivity and specificity vary depending on the choice of test and the test methods.
• The quality of studies assessing diagnostic test accuracy in TPMT testing was mixed, with thirty of fifty five studies categorized as high quality based on the quality appraisal.
• The low prevalence of patients with deficient TPMT activity or homogeneous TPMT mutations, the rarest mutations, made estimates of sensitivity of the tests for these mutations uncertain.
• The accuracy of genotyping is affected by the range of polymorphisms included in the test.
• Consistent testing protocols are needed to generate high quality evidence of clinical validity and clinical utility of TPMT genotyping technologies to ensure appropriate and consistent use in patient populations who would benefit from this testing.
and to appraise the quality of the evidence to identify the characteristics of high quality studies.

**Target Population**

This study sought to include evidence from all patient populations, age groups and clinical indications for which TPMT testing may be pursued.

**Methods**

**Systematic review**

The literature was systematically searched using eleven prominent electronic databases and a comprehensive list of grey literature sources. The detailed search strategy used did not limit to language or date, and included both methods of TPMT testing. Results of the search were reviewed to identify relevant studies meeting all inclusion criteria.

**Quality appraisal**

The Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS-2) was used to assess quality of the included studies. Data were extracted from each study included basic study design characteristics, study results, diagnostic test performance characteristics, and raw data to populate 2x2 and 3x3 contingency tables to calculate sensitivity, specificity, positive predictive value, negative predictive value, and concordance.

**Results**

**Systematic review**

The review process is summarized in Figure 1. Sixty-six studies met inclusion criteria and underwent quality appraisal. These 66 studies comprised three categories –phenotype-genotype comparisons, phenotype-phenotype comparisons and genotype-genotype comparisons.
Quality appraisal
In total, 30/55 phenotype-genotype comparisons were designated as high quality by the quality appraisal, and 6/11 phenotype-phenotype or genotype-genotype comparisons were designated as high quality. The appraisal characteristics for phenotype-genotype comparisons are presented in Table 1 and for phenotype-phenotype or genotype-genotype comparisons are presented in Table 2.

Studies considered of low quality generally contained unclear information relating to the quality components of the appraisal, as opposed to obvious bias or concerns for applicability. Thirteen of 30 high quality studies had low bias and low concern for applicability, while the remaining high quality studies had at least one domain with unclear or high risk associated with it. All of the high quality studies were published between 1997 and 2013, and examined a range of genotype and phenotype test methods.

Based on available data from 15 studies, the calculated sensitivity for genotyping to identify a homozygous mutation ranged from 0.0% to 100.0%. Based on data available from 26 studies, calculated specificity ranged from 97.8% to 100.0%. Based on available data from 25 studies, the calculated sensitivity to detect a homozygous or heterozygous mutation ranged from 13.4 to 100.0% and calculated specificity ranged from 90.9 to 100.0% using data from 26 studies.

Discussion

The choice of technologies available for the diagnosis of TPMT deficiency is varied. This review revealed a diverse and large body of literature assessing both phenotype and genotype technologies for TPMT testing across several disease states. There are limitations to both genotype testing and phenotype testing, and neither test can be referred to as the ‘gold standard’ for identifying TPMT deficiency.

The quality appraisal revealed that inadequate reporting of count data, descriptive information of index tests, reference tests, and recruitment methods, and study populations largely contributed to the exclusion of studies due to quality. Lack of reporting of diagnostic test accuracy indicates a need for guidance on reporting of test performance characteristics for diagnostic technologies. Thirty high quality studies comparing phenotype and genotype technologies were included in this
review. The number of polymorphisms included in genotype tests ranged from two to nine, with most studies including TPMT*2 and TPMT*3, the most common genetic variants in persons with deficient TPMT activity. Among the fifteen studies for which both sensitivity and specificity of genotyping could be calculated, ten demonstrated perfect (100%) sensitivity and specificity. The inference of perfect values is misleading, however. The low prevalence of homozygous mutations (0.3%) made it difficult for some study authors to generate sample sizes that were large enough for a stable rate of detection of homozygous mutations. The variation in sensitivity and specificity observed in the present review may also be related to the disease context. The tolerance for the risk of serious ADEs, and consequently values for sensitivity and specificity, may be different for chronic disease such as IBD and dermatological conditions versus life-threatening disease such as ALL.

Routine testing for all possible polymorphisms is more costly and unlikely to be feasible for health care institutions. Although current tests may become less costly in the future, there may also be mutations that have not yet been identified with current methods. Consideration also needs to be given to oversight and regulation, the applicability of pharmacogenetic discoveries in ethnically diverse populations and special populations (children, elderly), and to the anticipated shift from diagnostic pharmacogenetic testing to screening. Clinical and institutional decision-makers require high quality evidence of clinical validity and clinical utility of TPMT genotyping technologies to ensure appropriate and consistent use in patient populations who would benefit from this testing.
Figure 1. PRISMA flowchart

Records identified through citation database and grey literature searches, n = 4071

Duplicates removed n = 1983

Title and abstract screening n = 2088

Not eligible n = 1712
Unable to retrieve full text n = 3

Full text review n = 373

Not eligible n = 252
Insufficient data n = 55

Records for review n = 66
Table 1. QUADAS-2 results for high quality phenotype-genotype studies

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Domain 1 - Patient Selection</th>
<th>Domain 2 - Index Test</th>
<th>Domain 3 - Reference test</th>
<th>Domain 4 - Flow and Timing</th>
<th>Domain 5 - Genomics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author</td>
<td>Year</td>
<td>Risk of bias</td>
<td>Concern for applicability</td>
<td>Risk of bias</td>
<td>Concern for applicability</td>
<td>Risk of bias</td>
</tr>
<tr>
<td>Ben Salah ¹</td>
<td>2013</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Unclear</td>
</tr>
<tr>
<td>Fakhoury ²</td>
<td>2007</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Fangbin ³</td>
<td>2012</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Ford ⁴</td>
<td>2006</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Ford ⁵</td>
<td>2009</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Ganiere-Monteil ⁶</td>
<td>2004</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Gazouli ⁷</td>
<td>2012</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Hindorf ⁸</td>
<td>2012</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Unclear</td>
</tr>
<tr>
<td>Jorquera ⁹</td>
<td>2012</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Langley ¹⁰</td>
<td>2002</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Larussa ¹¹</td>
<td>2012</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Lennard ¹²</td>
<td>2012</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Liang ¹³</td>
<td>2013</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Loennechen ¹⁴</td>
<td>2001</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Ma ¹⁵</td>
<td>2006</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Unclear</td>
</tr>
<tr>
<td>Marinaki ¹⁶</td>
<td>2003</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Unclear</td>
<td>Unclear</td>
</tr>
<tr>
<td>Milek ¹⁷</td>
<td>2006</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Oselin ¹⁸</td>
<td>2006</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Domain 1 - Patient Selection</td>
<td>Domain 2 - Index Test</td>
<td>Domain 3 - Reference test</td>
<td>Domain 4 - Flow and Timing</td>
<td>Domain 5 - Genomics</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>------</td>
<td>------------------------------</td>
<td>-----------------------</td>
<td>--------------------------</td>
<td>---------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Risk of bias</td>
<td>Concern for applicability</td>
<td>Risk of bias</td>
<td>Concern for applicability</td>
<td>Risk of bias</td>
</tr>
<tr>
<td>Schaeffeler 19</td>
<td>2004</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Schwab 20</td>
<td>2002</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Serpe 21</td>
<td>2009</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Unclear</td>
<td>Low</td>
</tr>
<tr>
<td>Spire-Vayron de la Moureyre 22</td>
<td>1998</td>
<td>Unclear</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Spire-Vayron de la Moureyre 23</td>
<td>1998</td>
<td>Unclear</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>von Ahsen 24</td>
<td>2005</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Unclear</td>
<td>Unclear</td>
</tr>
<tr>
<td>Wennerstrand 25</td>
<td>2013</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Winter 26</td>
<td>2007</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Wusk 27</td>
<td>2004</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Xin 28</td>
<td>2009</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Yates 29</td>
<td>1997</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Zhang 30</td>
<td>2007</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Unclear</td>
<td>Low</td>
</tr>
</tbody>
</table>
Table 2. QUADAS-2 results for high quality phenotype-phenotype or genotype-genotype studies

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Domain 1 - Patient Selection</th>
<th>Domain 2 - Index Test</th>
<th>Domain 3 - Reference test</th>
<th>Domain 4 - Flow and Timing</th>
<th>Domain 5 - Genomics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Risk of bias</td>
<td>Concern for applicability</td>
<td>Risk of bias</td>
<td>Concern for applicability</td>
<td>Risk of bias</td>
</tr>
<tr>
<td>Chowdury 31</td>
<td>2007</td>
<td>Unclear</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Kim 32</td>
<td>2013</td>
<td>Unclear</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Lu 33</td>
<td>2005</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Unclear</td>
</tr>
<tr>
<td>Ma 34</td>
<td>2003</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Roman 35</td>
<td>2012</td>
<td>Unclear</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Schaeffeler 36</td>
<td>2008</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>
REFERENCES


