The Hospital for Sick Children
Technology Assessment at SickKids (TASK)

EXECUTIVE SUMMARY

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

Authors:

Lilla M. Roy, RN, BScN, MSc
Clinical Research Project Coordinator, Child Health Evaluative Services, The Hospital for Sick Children Peter Gilgan Centre for Research and Learning, Toronto

Wendy J. Ungar, MSc, PhD
Senior Scientist, Child Health Evaluative Sciences, The Hospital for Sick Children Peter Gilgan Centre for Research and Learning, Toronto; Professor, Health Policy, Management & Evaluation, University of Toronto

Richard M. Zur, PhD
Research Project Manager, Child Health Evaluative Services, The Hospital for Sick Children Peter Gilgan Centre for Research and Learning, Toronto

Corresponding Author:

Wendy J. Ungar, MSc, PhD
The Hospital for Sick Children Peter Gilgan Centre for Research and Learning
11th floor, 686 Bay Street
Toronto, ON, Canada M5G 0A4
tel: (416) 813-7654, extension 303487, fax: (416) 813-5979, e-mail: wendy.ungar@sickkids.ca
http://www.sickkids.ca/AboutSickKids/Directory/People/U/Wendy-Ungar.html

Report No. 2015-02

Date: July 29, 2015

Available at: http://lab.research.sickkids.ca/task/reports-theses/

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
Division Head, Clinical Pharmacology and Toxicology, The Hospital for Sick Children, Professor, Medicine, Pharmacology & Pharmacy, Department of Paediatrics, University of Toronto

Elizabeth Uleryk, MLS
Director, The Hospital for Sick Children Library, Toronto

James Whitlock, MD
Division Head/Chief Haematology/Oncology, The Hospital for Sick Children; Professor, 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.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to disclose.

EXECUTIVE SUMMARY

Introduction

Thiopurine S-methyltransferase (TPMT) is an enzyme that metabolizes thiopurine drugs which 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. There has long been phenotype blood testing to measure TPMT enzyme activity, and more recently a genotype test is sued to identify individuals with the genetic variants that determine TPMT activity. Uncertainty remains however, regarding which is the optimal test.

Objectives

The objectives of this study were to systematically review the literature on the performance characteristics of thiopurine testing for TPMT deficiency, to appraise the quality of the literature, and to identify the characteristics of high quality studies.

Methods

A systematic search of electronic databases was conducted, including Biosis, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Cochrane Database of Systematic Reviews (CDSR), Cochrane Central Register of Controlled Trials (CCTR), Database of Abstracts of Reviews of Effects (DARE), Health Technology Assessment (HTA), National Health Service Economic Evaluation Database (NHSEED), Embase, International Pharmaceutical Abstracts (IPA), Medline, and PubMed. Studies in any language comparing a genotype or phenotype technology to another genotype or phenotype technology were included. Studies must have been conducted in humans, and they must have reported (or provided data to calculate) sensitivity, specificity, negative predictive value (NPV), positive predictive value (PPV), or concordance between the two technologies.

The abstracts and full text of papers were reviewed to identify studies that met the inclusion criteria. The quality appraisal was completed using the Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS-2). Data extraction from the resulting studies included basic study design characteristics, study results, diagnostic test performance characteristics, and raw data

to populate 2x2 and 3x3 contingency tables to enable the calculation of sensitivity, specificity, PPV, NPV and concordance.

Results

Four thousand seventy-one studies were identified through the database and grey literature search. Three hundred and seventy three records required full text review, and 121 records were reviewed for relevant data. Sixty six studies had sufficient data for inclusion, and underwent quality appraisal. These 66 studies comprised three categories – a category of phenotype-genotype comparisons, and a category of phenotype-phenotype comparisons and genotype-genotype comparisons. In total, 30/55 phenotype-genotype comparisons were designated high quality by the quality appraisal, and 6/11 phenotype-phenotype or genotype-genotype comparisons were designated as high quality.

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% and with data that were available from 26 studies 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 specificity ranged from 90.9 to 100.0% using data available 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 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 preferred for chronic disease such as IBD and dermatological conditions versus life-threatening disease such as ALL.

Conclusion

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. This review of the literature comparing phenotype testing and genotype testing for TPMT status demonstrated a broad base of evidence these tests. The quality of the studies for assessing diagnostic test accuracy was mixed. The low prevalence of patients with deficient TPMT activity or homogeneous TPMT mutations made estimates of sensitivity of the tests uncertain. The accuracy of genotyping is also affected by the range of polymorphisms included in the test. Routine testing for all possible polymorphisms is more costly and unlikely to be feasible for health care institutions. Nevertheless, 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.