Article details: 2019-0084	
Title	A cost utility analysis of iron deficiency screening for infants at 18 months
Authors	Sarah Carsley PhD, Rui Fu MSc, Cornelia M. Borkhoff PhD, Nadine Reid PhD, Eva Baginska MBA, Catherine S. Birken MD MSc, Jonathon L. Maguire MD MSc, Rebecca Hancock-Howard PhD, Patricia C. Parkin MD, Peter C. Coyte PhD
Reviewer 1	James Ted McDonald
Institution	Department of Economics, University of New Brunswick, Fredericton, NB
General comments (author response in bold)	The authors note in the limitations section that estimates of the probability of poor functionality come from literature based on analyses in developing countries (e.g., Costa Rica, Indonesia and Turkey). It appears that these estimates relate to functional limitations for children with iron-deficient anemia (IDA) whereas what is being identified in the proposed screening is iron deficiency. **In this comment, the Reviewer highlights the distinction between ID and IDA. ID is an early stage which may persist, become chronic and at a late stage may be accompanied by anemia. According to principles of screening, it is the early, latent stage of a condition that should be targeted (rather than the late stage). In the 4th paragraph of Interpretation, we have carefully described the literature regarding poor outcomes of both ID and IDA. (Page 10, line 14 to Page 11, line 5)** 1. What is the evidence of functional delays, if any, for children who are ID but who do not progress to IDA? How likely is iron deficiency to progress to IDA in Canadian children if the ID is left untreated? **Given the known prevalence of ID of 10-15% (ie, approximately 12%) and prevalence of IDA of 2%, and the known physiology that the stage of ID always precedes IDA, we have calculated the conditional probability that a child progresses to IDA given that he or she has ID = 0.02/0.12 = 1 in 6. Therefore, we have revised the sentence in Introduction as follows: "In developed countries, the prevalence of ID in young children is approximately10-15%; and in approximately 1 in 6, ID may progress to anemia (iron deficiency anemia, IDA) with a prevalence of approximately 2%."** In the 4th paragraph of Interpretation we cite literature supporting evidence for an association between delays in children with ID (reference 14). Furthermore, we begin this paragraph with the statement: "The goal of screening for ID is early detection and treatment before progression to chronic, severe ID or IDA." To emphasize this point, we have revised this sentence to: **"In keeping with the principles published by the World Health Organization, the goal of screening for ID is early detection and treatment before progression to chronic, severe ID or IDA." Reference: Wilson JMG, Jungner G. Principles and Practice of Screening for Disease. Geneva, Switzerland: World Health Organization; 1968 (Page 1, line 6, Page 10, line 17)** It would seem reasonable to think that in typical Canadian diets the likelihood of progression to IDA would be lower than that observed in developing countries. Furthermore, the sample sizes in these studies are relatively small which raises issues of the generalizability of the results to a population setting. **We agree that the prevalence of ID/IDA is higher in developing countries than in developed countries. However, we have used prevalence from**

	developed countries, including a contemporary Canadian urban cohort. Furthermore we have conducted extensive sensitivity analysis in which we varied the prevalence of ID from 9.1% to 30.8% for the general population and from 11% to 51% for at-risk children. 2. The estimates rely on testing that is administered as part of a Well Baby Visit. Are there any data on the takeup of this screening by Ontario residents? I would expect that there would be selection effects based on socioeconomic status which would also possibly be correlated with ID. **The Ontario immunization schedule includes a well-baby visit at 18 months for DTaP-IPV-Hib (Diphtheria, Tetanus, Pertussis, Polio, Haemophilus influenzae type b). In 2015-2016, 84% of parents of 7-year old children reported they had received this vaccine (https://www.publichealthontario.ca/-/media/documents/immunization-coverage-2013-16.pdf?la=en) suggesting that a large number of children attend the 18-month visit.** **The reported uptake of the Ontario EWBV is 48%; however, this was determined using physician billing claims for this specific EWBV (which is an additional code which may be billed when physicians complete developmental assessment) and does not accurately reflect the uptake of the general 18-month visit. (Guttman A et al. Uptake of Ontario's Enhanced 18---Month Well---Baby Visit. Toronto, ON Institute for Clinical Evaluative Sciences; 2016.)** **We have revised the following sentence in paragraph 7 of Interpretation: "In the context of a primary care practice setting, screening during the 18 month visit is feasible due to the high uptake of this visit for immunizations and general health surveillance; and treatment of ID with diet advice and prescription of oral iron is feasible and utilizes the expertise of physicians and other members of the health care team." (Page 12, lines 6-10)** 3. Cost estimates for testing and treatment are based on marginal cost – the cost incurred in one extra child being tested and treated if indicated by the test results. However with an expansion of iron deficiency testing extended to the entire population of children aged ~18m, would additional costs system level costs be incurred? Does the current blood testing system have the excess capacity to accommodate such a marked expansion in demand for services that would be associated with universal testing? **This is a valid argument. However our decision tree model does not permit us to address potential system-level costs in the analysis. While we only considered patient-level marginal cost, we have done extensive sensitivity analysis to verify the impact of higher costs associated with screening on the cost-effectiveness results. For example, the total per-patient test cost in a universal screening program was varied from $28.48 to $37.03, a $8.55 (or 30%) marginal difference, which translates into a more than a one-million-dollar difference (using N=145214 babies born in Ontario in 2018). While the cost-effectiveness of a screening program is an initial step in health care policy, it is not the sole factor. A subsequent step would be to conduct a Budget Impact Analysis to determine the long-term system-level impact.** **We have revised our manuscript as follows: "There are several opportunities for future research. These include understanding the values and preferences of parents and practitioners for screening for ID; conducting a budget**

	impact analysis to address the system-level costs associated with screening;" (Page 13, lines 8-10) 4. The value of a DALY is sufficiently high relative to cost such that the test/treatment is cost effective by a wide margin, and sensitivity checks allowing for a range of values show robustness. Related literature shows similar results and the value of early childhood screening for a host of other conditions. Are there other obstacles, costs or barriers that are not included here that might explain the fact that more comprehensive screening of children is not already occurring? **At a practical level, fear of needles may be a barrier. However, parents likely place a higher value on their child's health and developmental outcomes. Those parents who place a higher value on avoiding needles and believe their child is at low risk for ID may prefer to decline screening. This is similar to any screening program. We have described this in our manuscript as follows (page 6): "A median utility of 1.0 has previously been identified for venipuncture, when parents were presented a scenario in which poor functional outcomes were possible; therefore, this was not included in the model.[37]"** **Economic analysis in child health is also a relatively underdeveloped field. As we discuss, citing the work of James Heckman, investment in child health and education provides a very high return on investment. However, as we have previously mentioned, cost-effectiveness studies are just the first step in evidence-based policy making.**
Reviewer 2	Amir Khakban
Institution	Collaboration for Outcomes Research and Evaluation, The University of British Columbia, Vancouver, BC
General comments (author response in bold)	Major Comments 1-It is not clear how authors connect information of quality adjusted life years of individual with iron deficiency to an individual with mild cognitive impairment. Authors assumed a healthy child has a utility score of 1 and lives as a normal individual with this utility rest of the life compare with a child suffering ID with utility score of .84 for lifetime. Using 0.16 utility score difference in QALYs of these two groups is a key assumption in this study. Nonetheless there is not sufficient evidence in the paper to support this assumption. **We agree that the utility screen difference in QALYS is a key assumption in our study. As our sensitivity analysis shows, when the utility difference is as low as 0.05, our results still show cost-effectiveness. Also, a utility score of 0.84 was assigned to children who developed poor functional outcomes, not necessarily to those who lived with ID. The probability of developing a poor functional outcome based on ID status was another input in our model. The editors have requested clarification on the utility assumptions through comments 8-13 above. We have responded to these comments. (Page 6, line 1; lines 5-8)** 2-The way that structure of model laid out and the result in its present form carries confusion in some of the outcomes. These are some of the concern with the results; On page 25 of 33, considering lowest QALYs of 0.84 for individual with impairment , highest QALYs 1 for healthy individual and life span of 80.5 year in the paper, it is not clear why average QALYs per person is calculated around 23.82 for the life

	span of individuals in three groups (table3). It should be between (80.5 * 0.84 and 80.5 * 1). **We choose 18-month after birth to be the time point at which we conduct this economic analysis. Hence, all costs and effects that occur in the future (beyond the 18th month) are discounted back to reflect their present values. Hence, the seemingly small effectiveness of no screening (23.82 years) is due to this discounting.** On Page 11of33 lines 40-54 numbers mentioned for screen negative and screen positive don't match with the reference table 3. "The cost of each screening strategy per child was $144.81 for universal, screen negative; $314.81 for universal, screen positive; $151.33 for targeted, screen negative; $321.33 for targeted, screen positive; and $0.00 for no screening." **In the text we reported the marginal cost (i.e., the per-patient cost) associated with each screening strategy. The costs reported in Table 3 are the average costs to the society associated with each screening strategy, discounted to present value (the 18th month after birth).** **We clarified in the text as follows: "The marginal cost per child of each screening strategy was $144.81 for universal, screen negative; $314.81 for universal, screen positive; $151.33 for targeted, screen negative; $321.33 for targeted, screen positive; and $0.00 for no screening." (Page 8, lines 3-6)** -The tornado diagram shows that lower specificity of the test decrease ICER. It shows with specificity of 0.495 (even lower than coin flip) test is still cost effective with lower ICER. **The tornado diagram is unable to present the direction of changes in ICER as each parameter is varied. Hence, although the ICER will increase as the test specificity decreases, this trend is not reflected in the tornado diagram.** Minor comments: 1-Author used a study as a reference for sensitivity (0.56) and specificity (0.99) of the test from 27 years ago (Guyatt et al., 1992). Is there any other method of screening with higher sensitivity available? **Although this article was published many years ago, this was a high quality synthesis of 55 studies. Subsequently serum ferritin has been considered to be the best test for diagnosis of iron deficiency. For children, it remains the test recommended by pediatric professional organizations such as the American Academy of Pediatrics and the World Health Organization. It is also the test now recommended by the IRON MOM pathway in obstetrics.** 2-On page 15 Line 1-20 the justification of QALY offset is not clear. **The QALY offset of 0.33 is due to the potential adverse event of gastrointestinal issues (e.g. constipation) that occurs with iron supplementation for 4 months (4/12 months). We have clarified in the table. (page 22)** 3-In the tornado diagram effect of lower and higher level of parameters on ICER is not clear. It is better to identify them with different color or put related number on the diagram. **We have addressed this comment in the new tornado diagram. (page 29)**