Canadian Agency for Drugs and Technologies in Health

RAPID RESPONSE REPORT:
SUMMARY WITH CRITICAL APPRAISAL

TITLE: **Newborn Screening for Cystic Fibrosis: A Review of the Clinical and Cost-Effectiveness**

DATE: 20 April 2012

CONTEXT AND POLICY ISSUES

Cystic fibrosis (CF) is an autosomal recessive disorder caused by mutations of the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) gene.[1,2] Although CF is found in most races, Caucasians have the highest incidence throughout the world.[1,3] According to the report of the Canadian Cystic Fibrosis Foundation Patient Data Registry,[4] in 2010, there were 3849 individuals with CF in Canada; Caucasians accounted for 93.5% of these cases.[4] CF results in abnormal trans-epithelial salt transport that interfere with the clearance of airway surface liquid. Clinically, CF is expressed by chronic airway infections leading to lung and respiratory damage.[5] Furthermore, patient with CF usually have pancreatic insufficiency and elevated sweat salt concentration.[1,5]

Sweat chloride test is considered the reference standard for CF diagnosis.[1,5-7] Sweat test results are considered as positive when the chloride (Cl^-) concentration is >60mmol/L, and negative for Cl^- concentration <30mmol/L[6] or <40mmol/L.[1] Borderline results (30-60 mmol/L) require an extended clinical and genetic evaluation to confirm the diagnosis. In practice, the sweat test is given for patients who are highly suspected to have CF; however, affected patients are often misdiagnosed because the majority of clinical symptoms are not specific to CF. Consequently, the sweat test is often delayed.[1]

Early recognition of CF through newborn screening (NPS) permits the allocation of preventive care for early respiratory and nutritional involvement.[3,8,9] As a result of these preventive measures and the improvement of patient care, the average life expectancy for CF patients in the USA increased from 27 years in 1985 to a median survival of 38.3 years in 2010.[10,11] Newborn screening was made possible by the discovery of the positive correlation between elevated immonureactive trypsinogen (IRT) levels in the blood of newborn infants affected with CF.[3,12]

Several CF NPS protocols are reported in the literature and are composed of one test or a combination of up-to three tests.[1] The following tests are frequently reported in the literature:

- IRT test is usually used as the first screening step.[1] It measures the IRT on a heel prick blood sample. High IRT values indicate an elevated risk of having CF; however, threshold value for the IRT test is dependent on the laboratory kits used, the population screened, and the screening protocols and algorithms employed.[1]

- DNA tests are used for the detection of CFTR gene mutations.[1,2] The number of the screened gene mutations varies from one mutation, mutation F508-del,[1] up to 36 mutations.[13]

- Pancreatic-associated protein (PAP) test.[1] PAP is a protein produced by a diseased pancreas and is elevated in the blood of CF patients.[13,14]

An effective screening protocol is expected to identify all affected infants (sensitivity approaching 100%); in the meanwhile, it should minimize the risk of disturbing the life of healthy babies and their families.[3] The current review evaluates the evidence on diagnostic accuracy and cost-effectiveness of newborn screening protocols for cystic fibrosis to inform policy decisions on adding this test to the screening panels already performed on newborns in Canada.

RESEARCH QUESTIONS

1. What is the evidence on the accuracy of newborn screening for cystic fibrosis?

2. What is the cost-effectiveness of newborn screening for cystic fibrosis?

KEY MESSAGE

Eight cystic fibrosis newborn screening protocols were identified in the reviewed literature. There was strong evidence of high sensitivity of the evaluated screening protocols and suggestive evidence of low positive predictive value (high false screen-positive). The cost of cystic fibrosis newborn screening was estimated for two protocols; the screening cost per newborn ranged from $US 4.48 to $US 6.78.

METHODS:

Literature Search Strategy

A limited literature search was conducted on key resources including PubMed, The Cochrane Library (2012, Issue 2), University of York Centre for Reviews and Dissemination (CRD) databases, Canadian and abbreviated list of major international health technology agencies, as well as a focused Internet search. No methodological filters were applied to limit retrieval by study type. Where possible, retrieval was limited to the human population. The search was also limited to English language documents published between January 1, 2007 and March 21, 2012.

Selection Criteria and Methods

One reviewer screened citations and selected studies. In the first level of screening, titles and abstracts were reviewed for relevance using a predefined checklist (Appendix 1). Full texts of

any relevant titles/abstracts were retrieved, and assessed based on the initial inclusion criteria (Appendix 2).

Table 1: Selection Criteria

Population	Newborns
Intervention	Any diagnostic test including sweat testing, genetic evaluation and/or extended clinical examination
Comparator	NA
Outcomes	Diagnosis accuracy of cystic fibrosis newborn screening (including sensitivity, specificity, positive and negative predictive values) Cost-effectiveness of new born screening for cystic fibrosis
Study Designs	Health technology assessments, systematic reviews, meta-analyses, randomized controlled trials, non-randomized controlled trials, diagnostic trials, retrospective controlled analysis, economic evaluations

Exclusion Criteria

Studies were excluded if they did not meet the selection criteria, provided the results of a descriptive study, or uncontrolled analytic studies (observational or experimental). Duplicate reports of the same trials were excluded along with studies included in the systematic review and health technology assessment. Case reports were also excluded.

Critical Appraisal of Individual Studies

Critical appraisal of the included studies was based on study design. The methodological quality of the included health technology assessment was evaluated using the "assessment of multiple systematic reviews" (AMSTAR).[15] AMSTAR is an 11-item checklist that has been developed to ensure reliability and construct validity of systematic reviews. On the other hand, the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool[16] was used to evaluate the included diagnostic studies. The QUADAS tool is a 14-item questionnaire that is used to evaluate bias, data variability, and quality of reporting in diagnostic studies. The methodological quality of the included cost-effectiveness studies were assessed using the guidelines for appraisal of economic studies by Drummond et al.[17] For the included studies a numeric score was not calculated. Instead, the strengths and limitations of the study were described.

There were no systematic reviews or randomized clinical trials identified for critical appraisal.

SUMMARY OF EVIDENCE

Quantity of Research Available

A total of 280 potential citations were identified by the search in bibliographic databases, with 257 citations being excluded during the title and abstract review based on irrelevance to the questions of interest. The full text documents of the remaining 23 articles were retrieved. One article was detected by the grey literature search. Of the 24 articles, 15 did not meet the eligibility criteria and were excluded; leaving 9 articles for this review.[1,13,14,18-23] A PRISMA diagram demonstrating the study selection process is presented in Appendix 3.

Summary of Study Characteristics

Nine articles that addressed at least one of the study questions were included in this review, including one health technology assessment (HTA),[1] seven primary diagnostic studies,[13,14,18-22], and one cost comparison study.[23] Details regarding primary studies characteristics are tabulated in Appendix 4.

The included HTA was prepared by the Institute of Health Economics in Alberta.[1] It provided the evidence of the diagnostic validity of CF NPS protocols; however, the objectives of the assessment were not clearly stated in the report. The assessment included systematic reviews and primary studies that compared two or more NBS protocols for CF and reported results in terms of diagnostic accuracy values (i.e., sensitivity, specificity, false positive/ negative rates, and positive/ negative predicative values) On the other hand, non-comparative studies that only reported the results of a single screening protocol were excluded from the assessment along with studies evaluating older patients (adolescents and adults). The assessment identified two systematic reviews and three primary studies not included in the systematic reviews.

Seven primary studies[13,14,18-22] published after the HTA were included in this review. Five studies evaluated the diagnostic validity of screening protocols.[13,20-22,24] Seven screening protocols, consisting of two or three screening tests, were evaluated. Tests included in these protocols were IRT/ IRT,[20,21] IRT/ DNA (single mutation),[20] IRT/ DNA (multiple mutations),[14,20,21], IRT/ PAP,[13,14], IRT/ DNA/ DNA-sequencing and IRT/PAP/ DNA/ DNA-sequencing;[13] and IRT/ IRT/ DNA screening tests.[22] Two trials used a fixed screen-positive threshold for IRT test that varied from >60µg/L[13] to >105 µg/L[21]; while floating threshold ranging from 96th percentile to 99.7th percentile for the other trials. The threshold used for PAP test varied from ≥1.0 µg/L[14] to ≥1.6 µg/L.[13] The number of genome mutations screened varied as well depending on the most prevalent mutations in each population evaluated. Details on the specifications of each protocol are provided in Appendix 4. Two studies evaluated the validity of a single screening test each.[18,19] One evaluated the diagnostic value of nasal potential difference in detecting and confirming CF when used in infants with hypertrypsinaemia and equivocal sweat test.[19] The second study compared the sweat test when measured by the nanoduct system to the sweat test measured by the quantitative pilocarpine iontophoresis chloride (QPIC).[18] In all included studies, sweat testing measured by the QIPC method was considered the reference test. The threshold used to confirm CF for the sweat test was ≥60 mmol/L and the border-line diagnosis was between 30 and 59 mmol/L. In the case of an equivocal sweat test, a confirmed diagnosis was based on extended clinical and genetic evaluations.

The included economic evaluation study was a cost comparison analysis based on simulated data using the decision-tree approach.[23] Data and assumptions used to run the simulation were based on a review of the literature. The analysis compared the costs of two screening protocols: the IRT/ IRT and IRT/ DNA protocols.

Summary of Critical Appraisal

The strengths and limitations of included studies are summarized in Appendix 5.

The HTA[1] included a comprehensive literature search of multiple databases and the grey literature. However, study selection and data extraction were done by one reviewer. The objectives of the review were not explicitly cited. Assessment of the scientific quality of the

included studies was not reported; furthermore, conclusions of the review did not consider the impact of studies' quality of their results.

In five of the diagnostic studies included in this review, the reference standard test (sweat test) was used for screen-positive infants.[13,14,20-22] Although screen-negative infants could be diagnosed clinically later on, false screen-negative results could not be analyzed or estimated in three studies.[13,14,22] Two studies included data from CF registries and surveillance systems in order to estimate the number of infants who had the disease but had CF screen-negative results.[20,21]

Two primary studies evaluated the diagnostic accuracy of two screening tests.[18,19] One trial included patients with hypertrypsinaemia; therefore, the results might not be representative of the general population who would receive the test.[19] In the other trial, a case-control study design was used; however, it was not reported if the execution, analysis and interpretation of the index test were done without the knowledge of the results of the reference standard.[18]

The economic evaluation clearly reported the objectives and study design.[23] The source of data used in the model was reported, and the assumptions were identified. The study compared the costs of two protocols; the first test of both of them was IRT test. However, the diagnostic threshold used for the IRT test differed in each protocol. Varying the threshold would logically have an effect on the estimates of diagnostic validity. Furthermore, the study included the cost of screen-negative results, but did not include as estimation for the cost of false screen-positive results. Positively-screening a healthy infant could affect the quality of life and the psychosocial relationship between the parents and their baby.[25,26]

Summary of Findings

What is the evidence on the accuracy of newborn screening for cystic fibrosis?

One HTA and seven primary studies addressed the accuracy of NPS in identifying CF. A summary of the study findings and authors' conclusions are provided in Appendix 6.

Health Technology Assessment

The Institute of Health Economics (IHE) in Alberta published an HTA in 2007 that included a systematic review of evidence published between 1996 and 2006.[1] From the included studies in the assessment, it was found that the sensitivity, specificity and positive predictive value were equivalent for the IRT/ IRT protocol and IRT/ DNA protocol, with prevalence rates of CF from 1/2000 to 1/4000. Furthermore, one included study in the assessment reported evidence that the IRT/ DNA / IRT protocol provided higher sensitivity and reduced false positive rate when compared with the IRT/ IRT protocol. The assessment included one study that evaluated the IRT/ PAP protocol. It was reported that the PAP test was less expensive and easier to implement than the DNA tests. However, the assessment reported that there was limited evidence on the diagnostic validity of the IRT/ PAP protocol. There was no economic evaluation of CF NBS in this HTA.

Primary studies

IRT/ IRT protocol
The IRT/ IRT protocol was evaluated in two trials and was compared with the same reference standard, the sweat test.[20,21] Sensitivity of the protocol varied from 80.2%[21] to 86.6%.[20] This variation might be explained by the different threshold used in each study; >105µg/L and >99.0th percentile, respectively. One study reported estimates of the specificity, positive predictive value, and negative predictive value of the protocol;[20] these were 99.4%, 3.5%, and 99.9% respectively. Both studies were retrospective analyses of registry data and could be potentially affected by reporting bias.

IRT/ DNA (single mutation) protocol
One study evaluated the diagnostic validity of the IRT/ DNA (single mutation) protocol.[20] The IRT threshold was >99th percentile and the genome mutation screened was F508-del. The protocol had equivalent sensitivity and specificity when compared to the IRT/ IRT protocol of the same trial (same threshold values and same population); 89.9% versus 86.6% and 99.9% versus 99.4% for the sensitivity and specificity of IRT/ DNA versus IRT/ IRT protocol. Nevertheless, the IRT/ DNA protocol had the advantage of using the same screening card for both the IRT and DNA tests; while for the IRT/ IRT protocol the infant had to be recalled for a second screening test, which involved extra administrative load and generated anxiety for the families.

IRT/ DNA (multiple mutations) protocol
This was the most common protocol and was evaluated in three studies.[14,20,21] A floating IRT threshold (>99.0th percentile) was used in two studies,[14,20] and the third study used both a floating threshold of >96th percentile and a fixed one of > 105µg/L. The number of genome mutations was 5,[14] 12,[20] and 25 mutations.[21] Screening sensitivity increased with the number of mutations screened; the highest sensitivity estimate was 96.2% in association with the IRT/ DNA (25 mutations),[21] and the lowest estimate was 71.4% in association with IRT/ DNA (5 mutations).[14] Specificity did not vary significantly between protocols; ranging from 99.8% to 99.9%.

IRT/ PAP protocol
The PAP test was evaluated in two studies as a second step after a positive IRT test.[13,14] The sensitivity of the protocol ranged from 85.7%[14] to 95.0%.[13] The reason for the higher sensitivity in the second case was due to the relatively low IRT threshold used (>60µg/L).[13] Specificity of the protocol was the same for both studies (99.9%).

IRT/ DNA/ DNA-sequencing and *IRT/ PAP/ DNA/ DNA-sequencing protocols*
One study evaluated both protocols.[13] DNA analysis screened for 32 mutations, and the DNA-sequencing of the CFTR gene was positive when two mutations had been identified. The study reported that there was no statistically significant difference between the protocols in their diagnostic validity.

IRT/ IRT/ DNA protocol
One study evaluated the protocol and used a floating IRT test threshold of >99.9th percentile.[22] The screening sensitivity was 99.7%, and it was the only estimate reported for the diagnostic validity of this protocol.

The Nasal Potential Difference (NPD) Test
One study evaluated the NPD in infants with hypertrypsinaemia for whom the sweat test was inconclusive.[19] Although estimates for the diagnostic validity were not reported, it was shown that higher NPD values were associated with CF diagnosed infants as compared to healthy controls.

Sweat test (Nanoduct method)
The nanoduct method of sweat testing was evaluated as an alternative to the quantitative pilocarpine iontophoresis test (QPIT), the reference standard for the sweat test.[18] It was reported that the nanoduct reduced the possibility of human error, stimulation time, and the required sample volume. The test provided a 100% sensitivity and 97.5% specificity. However, the use of this test was evaluated as an alternative procedure for the sweat test; therefore, it can only be used as confirmatory test and not as an initial screening test.

What is the cost-effectiveness of newborn screening for cystic fibrosis?

One cost comparison study was included and compared the cost of the IRT/IRT protocol with the IRT/DNA protocol.[23] The study was based on a simulation model based on the decision-tree approach. To run the model, the study used an estimated CF incidence in the USA of 1/4000 newborns, an estimated sensitivity of the IRT/ IRT protocol of 80%, and an estimated sensitivity of the IRT/ DNA protocol of 96%. Based on these estimations, it was reported that the IRT/ IRT protocol would offer an average overall cost savings of US $2.3 per newborn as compared to the IRT/DNA protocol; the total cost estimated for the IRT/DNA protocol was US $6.78 per newborn.

Limitations

None of the included studies in this review was conducted in a Canadian setting. Information about the incidence and prevalence of CF, the distribution of IRT test values, and the knowledge of the most common types of mutations associated with CF in the various ethnic subpopulations would be necessary to inform the implementation of newborn screening for CF.

Another limitation of this review is that the diagnostic validity of CF screening was evaluated without the clinical utility of CF screening. Evaluation of the clinical utility would inform policy makers about the benefits of adding CF NPS to the current screening panel. On the other hand, the economical evaluation was limited to the comparison of the costs of two protocols. This evaluation could be extended, along with the clinical utility analysis, in order to estimate the budget impact of CF screening versus no screening or clinical diagnosis.

CONCLUSIONS AND IMPLICATIONS FOR DECISION OR POLICY MAKING

Although the evaluated protocols were associated with high sensitivity and specificity, their positive predictive values were either low or not estimated. Therefore, evidence on the diagnostic validity of the evaluated screening protocol is inconclusive. The decision as to whether to add screening for CF in existing screening panels should consider the benefits and harms of the available screening protocols. Early detection and follow up of infants with CF might improve the clinical outcomes of the disease; however, harms might also be introduced with these protocols by the high number of false screen-positive results. The cost of adding CF screening should also be considered. The cost of CF newborn screening varied from US $ 4.48 per newborn for the IRT/ IRT protocol to US $6.78 per newborn with the IRT/ DNA protocol.

None of the reviewed studies evaluated the cost of the other available protocols. Since the goal of early screening of CF is to initiate early treatment and follow-up, the benefits of available therapeutic options may also be considered when deciding on whether to add CF screening to newborn screening panels.

PREPARED BY:
Canadian Agency for Drugs and Technologies in Health
Tel: 1-866-898-8439
www.cadth.ca

REFERENCES

1. Screening newborns for cystic fibrosis [Internet]. Edmonton: Institute of Health Economics; 2007 [cited 2012 Apr 9]. Available from: http://www.ihe.ca/documents/IHE_Report_Screening_Newborns_for_Cystic_Fibrosis_Feb_2007_0.pdf

2. Narzi L, Ferraguti G, Stamato A, Narzi F, Valentini SB, Lelli A, et al. Does cystic fibrosis neonatal screening detect atypical CF forms? Extended genetic characterization and 4-year clinical follow-up. Clin Genet. 2007 Jul;72(1):39-46.

3. Southern KW, Merelle MM, nkert-Roelse J, Nagelkerke AD. Newborn screening for cystic fibrosis [Internet]. Cochrane Database Syst Rev. 2009 [cited 2012 Apr 9];(1). Available from: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD001402.pub2/pdf

4. Cystic Fibrosis Canada. Canadian cystic fibrosis patient data registry report 2010 [Internet]. Toronto: Cystic Fibrosis Canada; 2010. [cited 2012 Apr 18]. Available from: http://www.cysticfibrosis.ca/assets/files/pdf/CPDR_ReportE.pdf

5. Jayaraj R, Barton PV, Newland P, Mountford R, Shaw NJ, McCarthy E, et al. A reference interval for sweat chloride in infants aged between five and six weeks of age. Ann Clin Biochem. 2009 Jan;46(Pt 1):73-8.

6. Soultan ZN, Foster MM, Newman NB, Anbar RD. Sweat chloride testing in infants identified as heterozygote carriers by newborn screening. J Pediatr. 2008 Dec;153(6):857-9.

7. Legrys VA, McColley SA, Li Z, Farrell PM. The need for quality improvement in sweat testing infants after newborn screening for cystic fibrosis. J Pediatr. 2010 Dec;157(6):1035-7.

8. Munck A, Dhondt JL, Sahler C, Roussey M. Implementation of the French nationwide cystic fibrosis newborn screening program. J Pediatr. 2008 Aug;153(2):228-33, 233.

9. Cystic Fibrosis Foundation, Borowitz D, Robinson KA, Rosenfeld M, Davis SD, Sabadosa KA, et al. Cystic Fibrosis Foundation evidence-based guidelines for management of infants with cystic fibrosis. J Pediatr. 2009 Dec;155(6 Suppl):S73-S93.

10. Cystic Fribrosis Foundation. Patient registry: annual data report 2010 [Internet]. Bethesda (MD): The Foundation; 2011[cited 2012 Apr 18]. Available from: http://www.cff.org/UploadedFiles/LivingWithCF/CareCenterNetwork/PatientRegistry/2010-Patient-Registry-Report.pdf

11. Briesacher BA, Quittner AL, Fouayzi H, Zhang J, Swensen A. Nationwide trends in the medical care costs of privately insured patients with cystic fibrosis (CF), 2001-2007. Pediatr Pulmonol. 2011 Aug;46(8):770-6.

12. Lindau-Shepard BA, Pass KA. Newborn screening for cystic fibrosis by use of a multiplex immunoassay. Clin Chem [Internet]. 2010 Mar [cited 2012 Mar 26];56(3):445-50. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3227416/pdf/nihms335100.pdf

13. Vernooij-van Langen AM, Loeber JG, Elvers B, Triepels RH, Gille JJ, van der Ploeg CP, et al. Novel strategies in newborn screening for cystic fibrosis: a prospective controlled study. Thorax. 2012 Jan 23.

14. Sommerburg O, Lindner M, Muckenthaler M, Kohlmueller D, Leible S, Feneberg R, et al. Initial evaluation of a biochemical cystic fibrosis newborn screening by sequential analysis of immunoreactive trypsinogen and pancreatitis-associated protein (IRT/PAP) as a strategy that does not involve DNA testing in a Northern European population. J Inherit Metab Dis. 2010 Oct;33(Suppl 2):S263-S271.

15. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol [Internet]. 2007 [cited 2012 Feb 24];7:10. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1810543/pdf/1471-2288-7-10.pdf

16. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol [Internet]. 2003 Nov 10 [cited 2012 Jan 19];3(25). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC305345

17. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. BMJ [Internet]. 1996 Aug 3 [cited 2012 Feb 23];313(7052):275-83. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2351717

18. Sands D, Oltarzewski M, Nowakowska A, Zybert K. Bilateral sweat tests with two different methods as a part of cystic fibrosis newborn screening (CF NBS) protocol and additional quality control. Folia Histochem Cytobiol. 2010 Sep 30;48(3):358-65.

19. Sermet-Gaudelus I, Girodon E, Roussel D, Deneuville E, Bui S, Huet F, et al. Measurement of nasal potential difference in young children with an equivocal sweat test following newborn screening for cystic fibrosis. Thorax. 2010 Jun;65(6):539-44.

20. Massie RJ, Curnow L, Glazner J, Armstrong DS, Francis I. Lessons learned from20 years of newborn screening for cystic fibrosis. Med J Aust [Internet]. 2012 Jan 16 [cited 2012 Mar 26];196(1):67-70. Available from: https://www.mja.com.au/sites/default/files/issues/196_01_160112/mas10686_fm.pdf

21. Kloosterboer M, Hoffman G, Rock M, Gershan W, Laxova A, Li Z, et al. Clarification of laboratory and clinical variables that influence cystic fibrosis newborn screening with initial analysis of immunoreactive trypsinogen. Pediatrics [Internet]. 2009 Feb [cited 2012 Mar 26];123(2):e338-e346. Available from: http://pediatrics.aappublications.org/content/123/2/e338.full.pdf+html

22. Sontag MK, Wright D, Beebe J, Accurso FJ, Sagel SD. A new cystic fibrosis newborn screening algorithm: IRT/IRT1 upward arrow/DNA. J Pediatr. 2009 Nov;155(5):618-22.

23. Wells J, Rosenberg M, Hoffman G, Anstead M, Farrell PM. A decision-tree approach to cost comparison of newborn screening strategies for cystic fibrosis. Pediatrics. 2012 Feb;129(2):e339-e347.

24. Collaco JM, Panny SR, Hamosh A, Mogayzel PJ, Jr. False negative cystic fibrosis newborn screen. Clin Pediatr (Phila). 2010 Mar;49(3):214-6.

25. Beucher J, Leray E, Deneuville E, Roblin M, Pin I, Bremont F, et al. Psychological effects of false-positive results in cystic fibrosis newborn screening: a two-year follow-up. J Pediatr. 2010 May;156(5):771-6, 776.

26. Tluczek A, Orland KM, Cavanagh L. Psychosocial consequences of false-positive newborn screens for cystic fibrosis. Qual Health Res. 2011 Feb;21(2):174-86.

Appendix 1: Title and abstract screening checklist

Reviewer: **Date:**

Ref ID: **First Author (year):**

1 What is the STUDY POPULATION in this article?	□ newborns (include) □ All other population (exclude)
2 What is the INTERVENTION?	□ Any analytical technique (include) □ Can't decide (include)
3 What is the TYPE OF STUDY reported in this article?	□ Report of a clinical trial (controlled; randomized/non-randomized) (include) □ Meta-analyses/systematic reviews/HTAs (include) □ Report of Diagnostic trial (include) □ Report of a controlled prospective or retrospective cohort study (include) □ Report of an analytical controlled cross-sectional study (include) □ Academic/narrative review, comment, editorial, letter, note, patient handout, study design description (exclude) □ All other study designs (exclude) □ Can't decide (include)
Selection decision:	□ **Include** □ **Exclude**

Appendix 2: Full text screening checklist

Reviewer: **Date:**

Ref ID: **First Author (year):**

1. **Did this article include newborns who underwent screening for cystic fibrosis***?*
- ☐ Yes (include)
- ☐ No (exclude)
- ☐ Maybe (include)

2. **Is the article the PRIMARY REPORT of the FINAL results from:**
- ☐ Report of a clinical trial (controlled/uncontrolled; randomized/non-randomized) (include)
- ☐ Meta-analyses/systematic reviews/HTAs (include)
- ☐ Report of a diagnostic trial (include)
- ☐ Report of a controlled prospective or retrospective cohort study (include)
- ☐ Report of a controlled cross-sectional study (include)
- ☐ All other study types (exclude)
- ☐ Can't decide (include)

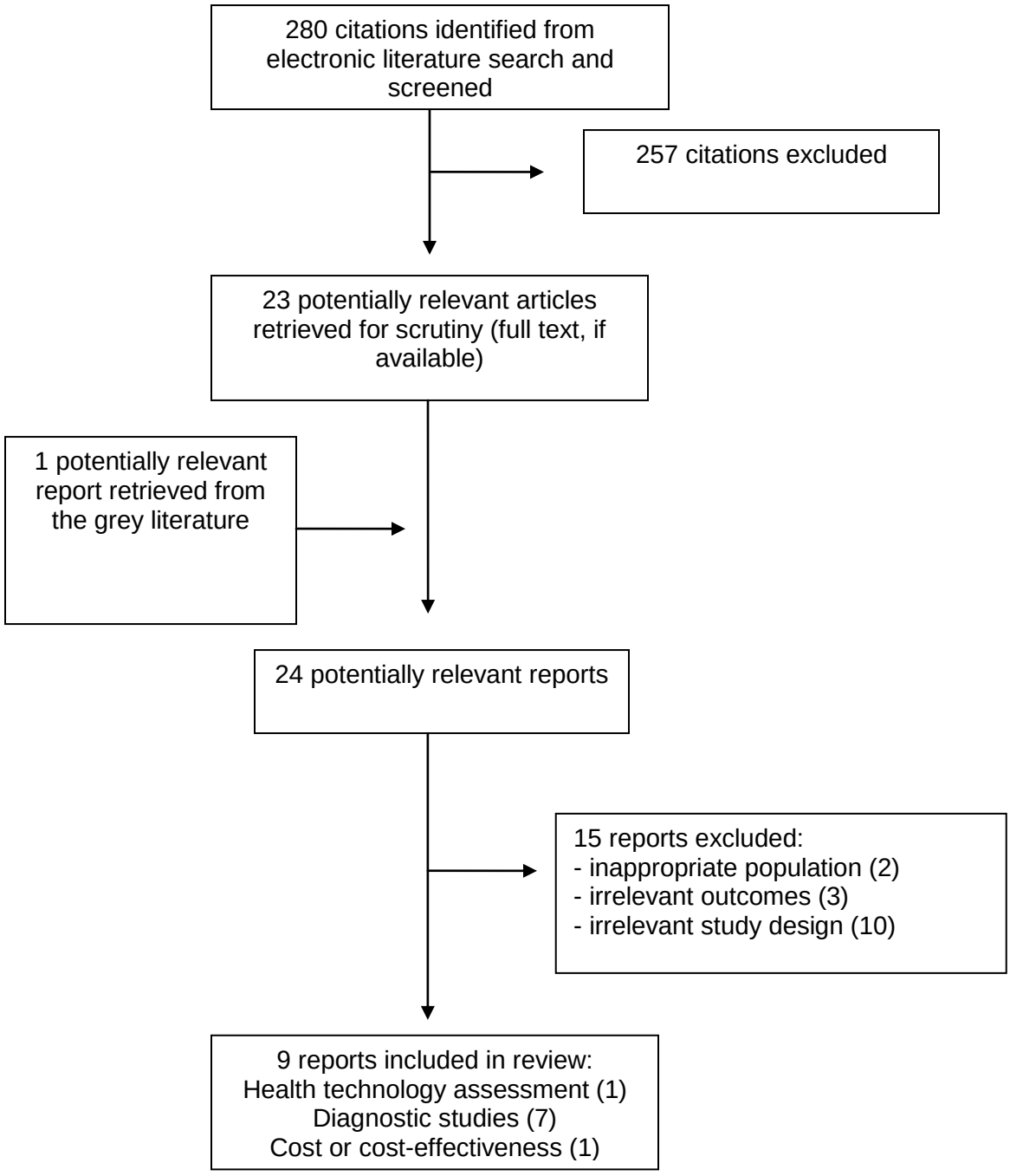
3. **What COMARATOR is used in the study?**
- ☐ Any comparator (include)
- ☐ No comparator (exclude)

4. **Include if the OUTCOME of interest in the study is one of the following:**
- ☐ Diagnostic accuracy of newborn screening (i.e. sensitivity, specificity, positive predictive value, negative predictive value) (include)
- ☐ Costs of newborn screening (include)
- ☐ Cost-effectiveness (e.g. cost per case detected, cost per quality adjusted life year, cost per life year) (include)
- ☐ None of the above (exclude)

5. **Final Decision**
- ☐ **Include**
- ☐ **Exclude**
- ☐ **Non-English /Unable to translate**

Reason for Exclusion:
- ☐ **Inappropriate study population**
- ☐ **Not study types of interest**
- ☐ **Not primary report of study**
- ☐ **Study description only**
- ☐ **No intervention of interest**
- ☐ **inappropriate control group**
- ☐ **No relevant outcomes**

Appendix 3: Selection of Included Studies



Appendix 4: Characteristics of the Included Primary Studies

First Author, Publication Year, Country	Study Design	Patient Characteristics (sample size)	Intervention	Comparator(s)	Clinical Outcomes
Diagnostic Studies* (Screening Protocols)					
Massie, 2012,[20] Australia	Retrospective diagnostic accuracy study	Newborns screened for CF in Victoria between 1989 and 2008. (n=1,318,615)	Three screening protocols were used in different time periods: • IRT/ IRT • IRT/ DNA (single mutation) • IRT/ DNA (12 mutations)	Sweat test and clinical assessment by a CF specialist	Diagnostic accuracy of the evaluated protocols in detecting CF
Vernooij-van Langen, 2012,[13] Netherlands	Prospective diagnostic accuracy study	Newborns screened for CF in Netherlands from 2008 and 2009. (n= 145,499)	Two screening protocols were evaluated: • IRT/ PAP • IRT/ DNA/ DNA sequencing IRT cut-off was ≥60µg/l and ≥ 1.6µg/l for PAP. DNA analysis screened for 36 mutations	Sweat test, CFTR mutations, and extended clinical evaluation	Diagnostic accuracy of the proposed protocols in detecting CF
Sommerburg, 2010,[14] Germany	Prospective diagnostic accuracy study	Newborns screened for CF in Heidelberg in a 20 months period. (n= 73,759)	Two screening protocols were evaluated: • IRT/ PAP (5 mutations) • IRT/ PAP cut-off values for IRT was >99.0th percentile and for PAP was ≥1.0 ng/ml.	Sweat test. If inconclusive sweat test infants had extended genetic and clinical examinations	Diagnostic accuracy of the proposed protocols in detecting CF
Sontag, 2009,[22] USA	Simulation study of the diagnostic accuracy based on retrospective data	Newborns screened for CF in Colorado between 1982 and 2007. (n >1,000,000)	Screening protocol composed of IRT/ IRT/ DNA (IRT cut-off 99.7% percentile)	Sweat test	Determination of an appropriate IRT cut-off Sensitivity of the proposed protocol in detecting CF
Kloosterboer, 2009,[21] USA	Retrospective diagnostic accuracy study	Newborns screened for CF in Wisconsin State between 1994. and 2004 (n= 660,443)	3 screening protocols were evaluated: • IRT (cut-off >105ng/mL)/ DNA (25 mutations) • IRT (cut-off 96th percentile)/ DNA (25 mutations) • IRT/IRT (cut-off >105ng/mL)	Sweat test and clinical confirmed CF as reported in the CF surveillance system	Factors that influence IRT concentrations. Sensitivity of the three screening protocols in detecting CF.
Diagnostic Studies* (Screening Tests)					
Sermet-Gaudelus, 2010,[19] France	Prospective case-control study	Infants with hypertrypsinaemia with an equivocal sweat screening test were referred from CF centres between 2006 and 2008. (n=23)	Nasal potential difference	Comprehensive mutation analyses and clinical assessment	Diagnostic value of nasal potential difference in confirming CF

First Author, Publication Year, Country	Study Design	Patient Characteristics (sample size)	Intervention	Comparator(s)	Clinical Outcomes
Sands, 2010,[18] Poland	Prospective diagnostic accuracy study	Infants with positive CF NBS who were referred for the sweat test at the CF centre in Warsaw between 2006 and 2009. Healthy infants were also included in the trial as controls. (n= 487 and 45 healthy control)	Sweat test measured by the Nanoduct system. Both conductivity and chloride concentration were evaluated by the new system	Sweat test measured by the quantitative pilocarpine iontophoresis chloride using the Schales and Schales mercuric nitrate procedure	Diagnostic accuracy of Nanoduct sweat test in the detection of CF

Cost-effectiveness Analysis				
	Study design	**Interventions**	**Data collection/ Assumption**	
Wells, 2012,[23] USA	Cost comparison – decision-tree approach	• IRT/ IRT (cut-off 105ng/mL and 70 ng/mL for the first and second IRT tests) • IRT/ DNA (23 mutations, and IRT cut-off 96th percentile)	• Incidence of CF in USA 1 in 4000 newborns • IRT sensitivity 80% • IRT/ DNA sensitivity 96%	Cost and cost saving simulation

CF= cystic fibrosis; CFTR= cystic fibrosis transmembrane conductance regulator; IRT= immonureactive trypsinogen; NBS= newborn screening; PAP= pancreas-associated protein
*diagnostic accuracy measurements included sensitivity, specificity, and positive and negative predictive values

Appendix 5: Critical Appraisal of Included Studies

First Author, Publication Year	Strengths	Limitations
Health Technology Assessment		
IHE, 2007[1]	• A comprehensive literature search was performed. • The characteristics of the included studies were provided.	• The objectives of the review were not stated in the report. • Literature search and data extraction was performed by one researcher. • The scientific quality assessment of the included studies was not documented. Furthermore, the review conclusion did not reflect the impact of studies quality on their results.
Diagnostic Studies (Screening Protocols)		
Massie, 2012[20]	• Patients were representative of those who would receive the test. • Screening protocols evaluated in the study reflected the clinical practice in Victoria at three different periods of time. • Newborns with screen-negative results were searched and captured through different CF databases	• The study was a retrospective analysis of multiple registry data. This type of analyses is vulnerable to reporting bias. • The reference test, sweat test, was performed based on the results of the index protocol and was not performed for all newborns. However, delayed clinical diagnosis was captured in the analysis; doing so would likely minimize this bias. • Different tests cut-offs were used reflecting the variations in clinical practice during the data collection period
Vernooij-van Langen, 2012[13]	• Patients were representative of those who would receive the test. • In the case of an equivocal diagnosis, infants had a regular follow-up at the CF centers during the first year of life.	• The reference standard (sweat test) was used for screen-positive individuals. The study duration and protocol did not allow for estimating the accuracy of the test in screen-negative individuals.
Sommerburg, 2010[14]	• The trial used one reference test, the sweat test. When equivocal results were obtained from the sweat test, infants had an extended work up including clinical evaluation, genetic testing and functional evaluation of CFTR-mediated Chloride secretion in rectal biopsy tissues.	• The reference standard (sweat test) was used for screen-positive individuals. The accuracy of the test in screen-negative individuals could not be captured in the trial; therefore the false negative estimation is likely to be biased. • The cut-value for the IRT test was based on the 99.9th percentile of the sample analyzed; however, the exact (actual) value was not reported.
Sontag, 2009[22]	• Patients were representative of those who would receive the test.	• The reference standard (sweat test) was used for screen-positive individuals. Infants with missed CF screening (based on the first tier in the protocol) were not captured in the trial; therefore the false negative estimation is likely to be biased. • The cut-value for the IRT test was based on the 99.9th percentile of the sample analyzed; however, the exact (actual) value was not reported. • Some of the tests and cut-off values

First Author, Publication Year	Strengths	Limitations
		used in the simulation model were hypothetical and may not reflect what is used in the clinical practice
Kloosterboer, 2009[21]	• Patients were representative of those who would receive the test. • To compensate for the false-negative results obtained for the screening test (protocol), the study used information from CF surveillance system available for the population under analysis.	• This was a retrospective analysis from a data registry, and the presence of reporting and recall bias was not evaluated in the analysis.
Diagnostic Studies (Screening Tests)		
Sermet-Gaudelus, 2010[19]	• All patients received the same diagnostic tests. • Execution of index tests was described in sufficient detail to permit replication of the test	• The trial enrolled infants with hypertrypsinaemia, and the results might not be representative of the general population who would receive the test. • The trial did not include infants with confirmed CF diagnosis (negative or positive); therefore, estimates of diagnostic accuracy could not be evaluated.
Sands, 2010[18]	• Patients were representative of those who would receive the test. • All patients received the same diagnostic tests.	• The study population consisted of individuals with confirmed diagnosis (CF positive or CF negative). It was not reported in the article if the execution, analysis and interpretation of the index test were done without knowledge of the results of the reference standard.
Cost-effectiveness Analysis		
Wells, 2012[23]	• The study design and objective were clearly reported. • The source of data used in the model was report, and the assumptions were identified.	• The study used different IRT cut-off values for each protocol in the comparison. A fixed cut-off was used for the IRT/ IRT protocol; while a floating IRT cut-off (96th percentile) was used for IRT/ DNA protocol. • The analysis included the cost of false screen-positive cases but did not include the impact of false-positive on the quality of life and psychological status of the infant and his family

CF= cystic fibrosis; CFTR= cystic fibrosis transmembrane conductance regulator; IRT= immonureactive trypsinogen; NBS= newborn screening; PAP= pancreas-associated protien

Appendix 6: Main Study Findings and Authors' Conclusions

First Author, Publication Year	Main Study Findings	Authors' Conclusions
Health Technology Assessment		
IHE, 2007[1]	• The review included two systematic reviews and three primary studies. • The two systematic reviews were consistent in that a single IRT test was associated with low positive predictive value. However, each systematic review had a different conclusion about the predictive value of the IRT/IRT and IRT/DNA protocols. • IRT/DNA (31 mutations)/IRT protocol was associated with higher sensitivity than IRT/IRT protocol. • IRT/DNA (16 mutations) protocol increased the diagnostic sensitivity but resulted in 26% more carrier identifications than the IRT/DNA (single mutation)* protocol. • IRT/PAP protocol would have equal or better sensitivity than that of the IRT/DNA (20 mutations) protocol.	Evidence about the diagnostic validity of each of the screening protocols was inconclusive. There were no Canadian studies included in the review, and the authors recognized that it is challenging to generalize the research evidence to the Canadian context.
Diagnostic Studies (Screening Protocols)		
Massie, 2012[20]	IRT cut-off value: >99th percentile Sweat test cut-off: ≥60mmol/L • **Sensitivity:** IRT/IRT: 86.6% IRT/DNA (single mutation)*: 89.9% IRT/DNA (12 mutations): 95.8% • **Specificity:** IRT/IRT: 99.4% IRT/DNA (single mutation)*:99.9% IRT/DNA (12 mutations):99.9% • **Positive predictive value:** IRT/IRT: 3.5% IRT/DNA (single mutation)*:20.1% IRT/DNA (12 mutations):18.3% • **Negative predictive value:** IRT/IRT: 99.9% IRT/DNA (single mutation)*: 99.9% IRT/DNA (12 mutations): 99.9%	*"Most babies with CF without meconium ileus, a family history or antenatal diagnosis are detected by newborn screening. Despite improved sensitivity with the 12-mutation analysis, most infants detected would have been diagnosed using the IRT/p.F508del protocol"*
Vernooij-van Langen, 2012[13]	IRT cut-off level ≥60µg/l PAP cut-off level ≥ 1.6µg/l (for IRT value ≥100µg/l) PAP cut-off level ≥30µg/l (for IRT value 60-100µg/l) sweat test cut-off: ≥60mmol/l • **Sensitivity:** IRT/PAP: 95.0% IRT/DNA (36 mutations)/ DNA sequencing: 100% IRT/PAP/ DNA (36 mutations)/ DNA sequencing: 95.0% • **Specificity:** IRT/PAP: 99.987% IRT/DNA (36 mutations)/ DNA sequencing: 99.991% IRT/PAP/ DNA (36 mutations)/ DNA sequencing: 99.998% • **Positive predictive value:** IRT/PAP: 12.3% IRT/DNA (36 mutations)/ DNA sequencing: 64.9%	*"All strategies performed well. Although there was no statistically significant difference in test performance, the IRT/DNA/sequencing strategy detected one infant that was missed by IRT/PAP (/DNA/sequencing). IRT/PAP may be the optimal choice if the use of DNA technology must be avoided. If identification of carriers and equivocal*

First Author, Publication Year	Main Study Findings	Authors' Conclusions
	IRT/PAP/ DNA (36 mutations)/ DNA sequencing: 87.5% • **Negative predictive value:** Not reported	*diagnosis is considered an important disadvantage, IRT/PAP/DNA/sequencing may be the best choice."*
Sommerburg, 2010[14]	IRT cut-off: >99.0th percentile PAP cut-off: ≥1.0 ng/ml Sweat test cut-off: >60mmol/L Sweat test boarder-line: 30-60mmol/L • **Sensitivity:** IRT/DNA (5 mutations): 71.4% IRT/PAP: 85.7% • **Specificity:** IRT/DNA (5 mutations): 99.9% IRT/PAP: 99.9% • **Positive predictive value:** IRT/DNA (5 mutations): 17.9% IRT/PAP: 12.2% • **Negative predictive value:** IRT/DNA (5 mutations): 99.9% IRT/PAP: 99.9%	*"Sequential measurement of IRT/PAP provides good sensitivity and specificity and allows reliable and cost-effective CF NBS which circumvents the necessity of genetic testing with its inherent ethical problems."*
Sontag, 2009[22]	IRT cut-off: >99.9th percentile Sweat test cut-off: ≥60mmol/L Sweat test boarder-line: 30-60mmol/L • **Sensitivity:** IRT/IRT/DNA: 99.7% • **Specificity:** Not reported • **Positive predictive value:** Not reported • **Negative predictive value:** Not reported	*"IRT/IRT1↑/DNA appears to improve cystic fibrosis newborn screen sensitivity while decreasing carrier identification, providing an alternative to IRT/IRT in states that obtain 2 blood spots."*
Kloosterboer, 2009[21]	Sweat test cut-off: not reported Sweat test boarder-line: not reported • **Sensitivity:** IRT (>105ng/mL)/ DNA (25 mutations): 90.6% IRT (96th percentile)/ DNA (25 mutations): 96.2% IRT/IRT (>105mg/mL): 80.2% • **Specificity:** IRT (96th percentile)/ DNA (25 mutations): 99.8% • **Positive predictive value:** Not reported • **Negative predictive value:** Not reported	*"Floating, rather than fixed, cut-off values for the initial IRT protein of any cystic fibrosis newborn screening protocol are generally necessary on the basis of the seasonal and reagent lot variations observed. Because of its lower sensitivity, IRT/IRT does not optimize detection of patients with cystic fibrosis."*
Diagnostic Studies		
Sermet-Gaudelus, 2010[19]	The nasal potential difference assessment was correlated with the CFTR genotype and was able to identify patients with CF • **Sensitivity:**	*"Evaluation of CFTR function in the nasal epithelium of young children with inconclusive results at*

First Author, Publication Year	Main Study Findings	Authors' Conclusions
	Not reported • **Specificity:** Not reported • **Positive predictive value:** Not reported • **Negative predictive value:** Not reported	*CF newborn screening is a usefull diagnostic tool for CF."*
Sands, 2010[18]	487 infants who had positive screening tests (IRT>99.4 percentile and one or two CFTR mutations) were evaluated with the sweat test (reference test: QPIT method; index test: Nanoduct method) Values of the Nanoduct methods: - Conductivity cut-off: 50 mmol/L - Chloride (Cl^-) concentration cut-off: 34 mmol/L • **Sensitivity:** Sweat test (Nanoduct method – conductivity): 100.0% Sweat test (Nanoduct method – Cl^- concentration): 100.0% • **Specificity:** Sweat test (Nanoduct method – conductivity): 97.5% Sweat test (Nanoduct method – Cl^- concentration): 97.5% • **Positive predictive value:** Sweat test (Nanoduct method – conductivity): 79.25% Sweat test (Nanoduct method – Cl^- concentration): 80.0% • **Negative predictive value:** Sweat test (Nanoduct method – conductivity): 100.0% Sweat test (Nanoduct method – Cl^- concentration): 100.0%	*"Nanoduct is a very useful and reliable tool in CF NBS protocol, allowing more time efficient organization of the diagnostic and training procedures. Simultaneous bilateral sweat testing with two different methods (concentration and conductivity) provides an extra quality control system."*
Cost-effectiveness Analysis		
Wells, 2012[23]	• Substantial number of potential missed diagnosis for IRT/ IRT system (10/100,000) versus IRT/ DNA (2.9/100,000). • The IRT/ IRT protocol offered an average overall cost savings of US $2.3 per newborn; the total cost estimated for IRT/ DNA protocol was US $6.78 per newborn.	"*The IRT/IRT screening algorithm reduces the costs to laboratories and insurance companies but has more system failures. IRT/DNA offers other advantages, including fewer delayed diagnoses and lower out-of pocket costs to families."*

CF= cystic fibrosis; IRT= immune-reactive trypsinogen; QPIT= quantitative pilocarpine iontophoresis test; PAP= pancreatitis-associated protein; CFTR= cystic fibrosis trans-membrane; NBS= newborn screening;

* Single mutation screening included the detection of ΔF508del mutation