World Health Organization

WHO/HIV/2014.37

Global Hepatitis Programme

Guideline development for Hepatitis C virus Screening, Care and Treatment in low- and middle-income countries
PICO 4 Care (Fibrosis) – Decision Making Table

Health system and public health evidence to recommendations framework

Population: People living with chronic HCV infection being assessed for HCV therapy
Intervention: Fibrosis stage determined by: liver biopsy, transient elastography, APRI, Forns, FIB4, and Fibrotest.
Comparison: Fibrosis stage determined by liver biopsy.
Outcomes:

1. Sensitivity/Specificity to detect F0-1 vs. F2-3-4 and F0-1-2-3 vs. F4
2. Cost/Cost-effectiveness

Background: The spectrum of disease in those infected with HCV extends from mild fibrosis to cirrhosis and hepatocellular carcinoma. Compensated cirrhosis may progress over time to decompensated cirrhosis associated with ascites, oesophageal and gastric varices and eventually to liver failure, renal failure and sepsis, all of which are life-threatening. Treatment of HCV must be commenced prior to the onset of decompensated disease as it may precipitate liver failure and death if administered at this stage. The diagnosis of decompensated liver disease is clinical and therefore a careful medical examination of patients must be made prior to commencing therapy. While early treatment of HCV is more efficacious than late treatment, resource constraints in low and middle income settings may necessitate treatment at a later stage. In order to prioritise treatment, a number of systems may be used to assess fibrosis stage. Traditionally, the liver biopsy is used to stage disease but use of this test is limited by the possibility of complications and the need for careful histological interpretation. A number of non-invasive scoring systems based on serum testing (APRI, Forns, Fibrotest, FIB4) and transient elastography were evaluated against liver biopsy as the gold standard.

	CRITERIA	JUDGEMENTS	RESEARCH EVIDENCE	ADDITIONAL INFORMATION
PROBLEM	**Is the problem a priority?**	*No* ☐ *Probably No* ☐ *Uncertain* ☐ *Probably Yes* ☐ *Yes* ☑ ***Varies*** ☐	HCV is a prevalent disease with substantial associated health risks. Patients may be assessed in order to stage disease with the aim if prioritization of those with more advanced disease (>F2 or F4 disease) in resource-limited settings..	
	Are a large number of people affected?	*No* ☐ *Probably No* ☐ *Uncertain* ☐ *Probably Yes* ☐ *Yes* ☑ ***Varies*** ☐	130-170 million people are affected, mostly in resource-limited settings.	

BENEFITS & HARMS OF THE OPTIONS

CRITERIA	JUDGEMENTS	RESEARCH EVIDENCE	ADDITIONAL INFORMATION
Are the desirable anticipated effects large?	*No* ☐ *Probably No* ☐ *Uncertain* ☐ *Probably Yes* ☐ *Yes* ☑ ***Varies*** ☐	A systematic review evaluating the sensitivity and specificity, risks and benefits and costs of non-invasive tests was carried out. APRI, FIB-4 and Forns have dual cut-offs for diagnosing specific fibrosis stages: a high cut-off with high specificity and a low cut-off with high sensitivity. Transient elastography has single cut-offs. The summary sensitivity and specificity for the detection of significant fibrosis (≥F2 stage) ranged from 39-96% depending on the test and cut-off value used. Using a low cut-off, transient elastography, APRI, Forns and FIB4 performed well with a sensitivity of ≥79% (79%, 82%, 88% and 89% respectively). Specificity ranged from 42-96% with transient elastography, APRI and Forns (high cut-off values) giving the highest results (83%, 92% and 96% respectively). The sensitivity and specificity of each test were evaluated for the detection of cirrhosis (F4 stage). The FIB4 and Forns tests were not included in this evaluation, as these tests were specifically developed for the diagnosis of F2. Fibroscan and APRI (low cut-off) gave the highest sensitivity results (89% and 77% respectively) and also the highest specificity results (high cut-off for APRI 94%, transient elastography 91%). The impact of using each test in two different prevalence settings (pre-test probabilities of 20% and 80% for ≥F2, 5% and 40% for cirrhosis) was considered for APRI versus transient elastography in a cohort of 1000 hypothetical patients. In order to reduce the cost impact of false positive results in low disease prevalence settings (5% cirrhosis or 20% prevalence of infection), the GDG considered that the APRI high cut off (1.0) would be preferable to the APRI low cut-off. In higher disease prevalence settings, (40% prevalence of cirrhosis or 80% prevalence of disease), the APRI low cut off (0.5) is preferable to avoid false negative results. A number of caveats were considered. Firstly, the APRI scoring system is less reliable in patients with HIV due to the possibility of throbocytopaenia associated with infection rather than cirrhosis. Theoretically, the FIB4 test could also be affected by thrombocytopaenia but this scoring system was first evaluated in patients with HIV and was found to perform well. Transient elastography may be artificially increased by a number of factors, including acute inflammation, elevated portal pressure, liver congestion (e.g. cardiac failure), a recent meal, amyloidosis and cholestasis. Link to systematic review	
Are the undesirable anticipated effects small?	*No* ☐ *Probably No* ☐ *Uncertain* ☐ *Probably Yes* ☐ *Yes* ☑ ***Varies*** ☐		
What is the overall certainty of this evidence?	*No included studies* ☐ *Very low* ☐ *Low* ☑ *Moderate* ☐ *High* ☐		

Problem: [Problem]

Option: [Option]

Comparison: [Comparison]

Setting: [Setting]

	CRITERIA	JUDGEMENTS	RESEARCH EVIDENCE	ADDITIONAL INFORMATION
VALUES	**How certain is the relative importance of the desirable and undesirable outcomes?**	*Important uncertainty or variability* ☐ *Possibly important uncertainty or variability* ☐ *Probably no important uncertainty or variability* ☑ *No important uncertainty or variability* ☐ *No known undesirable outcomes* ☐	Desirable outcomes were considered by the committee to be identification of patients to be prioritised for treatment, the lack of side effects associated with non-invasive testing and the availability of non-invasive testing in a variety of resource limited settings.Undesirable outcomes included the posibility of false positive case identification (increasing resource use and the potential for side effects associated with therapy) and false negative case identification (resulting in the possibility of death from cirrhosis and hepatocellular carcinoma. Overall, the potential increase in treatment availabiliyt resulting from increased access to non-invasive monitoring and a reduction in side effects was felt to outweigh the potential harms of false positive and negative case identification.	
	Are the desirable effects large relative to undesirable effects?	*No* ☐ *Probably No* ☐ *Uncertain* ☐ *Probably Yes* ☑ *Yes* ☐ ***Varies*** ☐		

	CRITERIA	JUDGEMENTS	RESEARCH EVIDENCE	ADDITIONAL INFORMATION
RESOURCE USE	**Are the resources required small?**	*No* ☐ *Probably No* ☐ *Uncertain* ☐ *Probably Yes* ☐ *Yes* ☐ ***Varies*** ☑	***Main resource requirements*** **Resource use** A presentation by Louise Longworth was considered by the committee. In the UK, APRI, FIB4 and Forns cost approximately £3 each, while the Fibrotest costs £43. Non-invasive imaging can be used out to assess liver fibrosis using transient elastography (Fibroscan). Resource use was discussed by the GDG committee and felt likely to be a major constraint in low and middle income settings; the APRI test was considered in most settings to be cheaper but in other settings, transient elastography or other measures of fibrosis could be employed if available.	
	Is the incremental cost small relative to the net benefits?	*No* ☐ *Probably No* ☐ *Uncertain* ☐ *Probably Yes* ☐ *Yes* ☐ ***Varies*** ☑		
EQUITY	**What would be the impact on health inequities?**	*Increased* ☐ *Probably increased* ☐ *Uncertain* ☐ *Probably reduced* ☑ *Reduced* ☐ ***Varies*** ☐	Increased availability of disease staging and as a result treatment was considered to be likely to improve health inequities.	
ACCEPTABILITY	**Is the option acceptable to key stakeholders?**	*No* ☐ *Probably No* ☐ *Uncertain* ☐ *Probably Yes* ☑ *Yes* ☐ ***Varies*** ☐	The option was considered likely to be acceptable to key stakeholders.	
FEASIBILITY	**Is the option feasible to implement?**	*No* ☐ *Probably No* ☐ *Uncertain* ☐ *Probably Yes* ☑ *Yes* ☐ ***Varies*** ☐	APRI and FIB4 were considered to be feasible as the tests included in these are required for treatment monitoring. Forns and Fibrotest were considered to be less feasible to implement. Transient elastography was available in some settings for example in some centres in Egypt, but not in others.	

Balance of consequences	Undesirable consequences *clearly outweigh* desirable consequences in most settings	Undesirable consequences *probably outweigh* desirable consequences in most settings	The balance between desirable and undesirable consequences *is closely balanced or uncertain*	Desirable consequences *probably outweigh* undesirable consequences in most settings	Desirable consequences *clearly outweigh* undesirable consequences in most settings
	☐	☐	☐	☑	☐

Type of recommendation	We recommend against the option	We suggest considering the option	We recommend the option
	☐	☐ Only in the context of rigorous research ☐ Only with targeted monitoring and evaluation ☐ Only in specific contexts	☑

Recommendation

- For assessment of hepatic fibrosis, it is suggested that simple non-invasive tests such as APRI and FIB4 be used rather than other non-invasive tests that require more resource such as elastography or Fibrotest.
- Conditional recommendation, low quality of evidence

Justification

The recommendation was based on the likely availability of testing in resource limited settings; the APRI and FIB4 scoring systems are based on tests that would be required for monitoring of patients on therapy and therefore do not incur additional cost. Other scoring systems or transient elastography may also be appropriate in different settings depending on resource constraints and local pricing.

Implementation considerations

Policy decisions on what stage of disease to treat at will depend on the local availability of treatment and cost-effectiveness evaluations.

Monitoring and evaluation

Research priorities

Research is required on the use of non-invasive testing in low and middle income settings

Evidence profile [title]

Authors: David Hunt, Esther Aspinall, and Hamish Innes
Date: 2013-05-16
Question:
Settings: Individuals with chronic HCV infection
Bibliography: [Citation text]

Table 1: Summary results for patients with >F2 disease

Test	Studies,n	Patients	Se	95% CI	Sp	95% CI
APRI_low	47	11696	**82**	77-86	**57**	49-65
APRI_high	36	9602	**39**	32-47	**92**	89-94
Fibroscan	37	8346	**79**	74-84	**83**	77-88
Fibrotest	17	5083	**68**	57-77	**72**	67-77
FIB4_low	11	2744	**89**	79-95	**42**	25-61
FIB4_high	9	2115	**59**	43-73	**74**	56-87
Forns_low	18	4747	**88**	83-91	**40**	33-48
Forns_high	15	4132	**35**	29-41	**96**	92-98
PLT	10	2849	**50**	41-59	**89**	83-93
AST_ALT ratio	7	1665	**44**	27-63	**71**	62-78

Table 2: Summary results for patients with F4 disease

Test	Studies, n	Patients	Se	95% CI	Sp	95% CI
APRI_low	24	7301	**0.77**	0.73-0.81	**0.78**	0.74-0.81

APRI_high	19	6930	**0.48**	0.41-0.56	**0.94**	0.91-0.95
Fibroscan	36	7923	**0.89**	0.84-0.92	**0.91**	0.89-0.93
Fibrotest	8	3724	**0.60**	0.43-0.76	**0.86**	0.81-0.91
PLT	13	2861	**0.49**	0.39-0.59	**0.87**	0.75-0.94
AST_ALT ratio	10	1984	**0.68**	0.59-0.76	**0.86**	0.72-0.94

Table 3A: TE vs. APRI high cut-off for >F2

Test Important Outcome	Results per 1000 patients tested (95% CI)			
	Pre-test probability 20%		Pre-test probability 80%	
	Elasto-graphy	APRI	Elasto-graphy	APRI
True Positive (TP)	158 (148-168)	78 (64-94)	632 (592-672)	312 (256-376)
TP absolute difference[2]	80 more (74-84 more)		320 more (296-336 more)	
False Positive (FP)	136 (96-184)	64 (48-88)	34 (24-46)	16 (12-22)
FP absolute difference[2]	72 more (48-96 more)		18 more (12-24 more)	
False Negative (FN)	42 (32-52)	122 (106-136)	168 (128-208)	488 (424-544)
FN absolute difference[2]	80 fewer (54-104 fewer)		320 fewer (296-336 fewer)	
True Negative (FN)	664 (616-704)	736 (712-752)	166 (154-176)	184 (178-188)
TN absolute difference[2]	72 fewer (48-96 fewer)		18 fewer (12-24 fewer)	

Table 3B: TE vs. APRI low cut-off for >F2

Test Important Outcome	Results per 1000 patients tested (95% CI)			
	Pre-test probability 20%		Pre-test probability 80%	
	Elasto-graphy	APRI	Elasto-graphy	APRI
True Positive (TP)	158 (148-168)	164 (154-172)	632 (592-672)	656 (616-688)
TP absolute difference[2]	6 fewer (4-6 fewer)		24 fewer (16-24 fewer)	
False Positive (FP)	136 (96-184)	344 (280-408)	34 (24-46)	86 (70-102)
FP absolute difference[2]	208 fewer (184-224 fewer)		52 fewer (46-56 fewer)	
False Negative (FN)	42 (32-52)	36 (28-46)	168 (128-208)	144 (112-184)
FN absolute difference[2]	6 more (4-6 more)		24 more (16-24 more)	
True Negative (FN)	664 (616-704)	456 (392-520)	166 (154-176)	114 (98-130)
TN absolute difference[2]	208 more (184-224 more)		52 more (46-56 more)	

Table 4A: TE vs. APRI low cut-off for F4

Test Important Outcome	Results per 1000 patients tested (95% CI)			
	Pre-test probability 5%		Pre-test probability 40%	
	Elasto-graphy	APRI	Elasto-graphy	APRI
True Positive (TP)	45 (42-46)	39 (37-41)	356 (336-368)	308 (292-324)
TP absolute difference[2]	6 more (1-9 more)		48 more (12-76 more)	
False Positive (FP)	86 (67-105)	209 (181-247)	54 (42-66)	132 (114-156
FP absolute difference[2]	123 fewer (114-142 fewer)		78 fewer (72-90 fewer)	
False Negative (FN)	6 (4-8)	12 (10-14)	44 (32-64)	92 (76-108)
FN absolute difference[2]	6 fewer (2-10 fewer)		48 fewer (12-76 fewer)	
True Negative (FN)	865 (846-884)	741 (703-770)	546 (534-558)	468 (444-486)
TN absolute difference[2]	124 more (114-143 more)		78 more (72-90 more)	

Table 4B: TE vs. APRI high cut-off for F4

Test Important Outcome	Results per 1000 patients tested (95% CI)			
	Pre-test probability 5%		Pre-test probability 40%	
	Elasto-graphy	APRI	Elasto-graphy	APRI
True Positive (TP)	45 (42-46)	24 (21-28)	356 (336-368)	192 (164-224)
TP absolute difference[2]	21 more (14-25 more)		164 more (144-172 more)	
False Positive (FP)	86 (67-105)	57 (47-86)	54 (42-66)	36 (30-54)
FP absolute difference[2]	29 more (19-58 more)		18 more (12-36 more)	
False Negative (FN)	6 (4-8)	26 (22-30)	44 (32-64)	208 (176-236)
FN absolute difference[2]	20 fewer (18-22 fewer)		164 fewer (144-172 fewer)	
True Negative (FN)	865 (846-884)	893 (865-903)	546 (534-558)	564 (546-570)
TN absolute difference[2]	28 fewer (21-57 fewer)		18 fewer (12-36 fewer)	

Problem: [Problem] ***Option:*** [Option] ***Comparison:*** [Comparison] ***Setting:*** [Setting]

References

Definitions for ratings of the certainty of the evidence (GRADE)**

Ratings	Definitions	Implications
⊕⊕⊕⊕ High	This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different* is low.	This evidence provides a very good basis for making a decision about whether to implement the intervention. Impact evaluation and monitoring of the impact are unlikely to be needed if it is implemented.
⊕⊕⊕◯ Moderate	This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different[4] is moderate.	This evidence provides a good basis for making a decision about whether to implement the intervention. Monitoring of the impact is likely to be needed and impact evaluation may be warranted if it is implemented.
⊕⊕◯◯ Low	This research provides some indication of the likely effect. However, the likelihood that it will be substantially different[4] is high.	This evidence provides some basis for making a decision about whether to implement the intervention. Impact evaluation is likely to be warranted if it is implemented.
⊕◯◯◯ Very low	This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different[4] is very high.	This evidence does not provide a good basis for making a decision about whether to implement the intervention. Impact evaluation is very likely to be warranted if it is implemented.

*Substantially different: large enough difference that it might have an effect on a decision

**The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group began in the year 2000 as an informal collaboration of people with an interest in addressing the shortcomings of present grading systems in health care. The working group has developed a common, sensible and transparent approach to grading quality of evidence and strength of recommendations. Many international organizations have provided input into the development of the approach and have started using it.

(Return)

For most recent version of this framework (and additional frameworks): www.decide-collaboration.eu/WP5/Strategies/Framework