

Consistency of Browser-Based Accessibility Evaluation Tools: An Empirical Study Across EU Healthcare Websites Using a Weighted Accessibility Metric*

Katarzyna MACIEJEWSKA

Polish-Japanese Academy of Information Technology (PJATK),
Department of Software Engineering, Warsaw, Poland,

Correspondence should be addressed to: Katarzyna MACIEJEWSKA, kata@pjwstk.edu.pl

*Presented at the 47th IBIMA International Conference, 29-30 June 2026, Madrid, Spain

Abstract

Automated browser-based tools are widely used to evaluate web accessibility compliance with the Web Content Accessibility Guidelines (WCAG). However, these tools use different underlying engines and rule sets, and it remains unclear whether they produce consistent results when applied to the same websites. This study compares three browser-based automated accessibility evaluation tools, WAVE, Accessibility Insights for Web, and QualWeb, applied to the homepages of 20 healthcare-related websites from five European Union countries. The countries were selected to represent a range of accessibility maturity based on the Digital Economy and Society Index (DESI, European Commission, 2024). To enable systematic cross-tool comparison, this study proposes the Weighted Accessibility Concern Index (WACI), a weighted composite metric that aggregates detected errors, contrast failures, and alerts by severity. The absence of a standardised output format across tools was a key finding of this study. It also motivated the development of replicable extraction procedures that enabled the application of a single WACI formula across all three tools. Results indicate moderate but statistically significant rank agreement between tools (Kendall's $\tau = 0.343\text{--}0.516$, all $p < 0.05$), while substantial differences in absolute WACI scores were observed. Sweden and Finland achieved the lowest mean WACI scores, Romania recorded the highest burden across all tools. Germany showed unexpectedly high scores despite its digital maturity ranking. The study contributes a replicable methodology for browser-based tool comparison in a European regulatory context, following the entry into force of the European Accessibility Act, whose accessibility requirements apply from June 2025.

Keywords: web accessibility, WCAG, WACI, accessibility evaluation tools

Introduction

Despite growing legal obligations, research consistently shows that most websites fail basic accessibility requirements. The WebAIM Million 2025 report found that 95.9% of the top one million homepages contained detectable WCAG errors. Automated browser-based tools are widely used to evaluate accessibility at scale, but

they use different engines, rule sets, and severity classifications, raising the question of whether they produce consistent results when applied to the same pages.

This study focuses specifically on browser-based tools, as distinct from command-line or API-based tools such as `Pally` or `axe-core CLI`. A separate comparison of command-line tools applied to the same sample is planned as a follow-up study. The present study examines three browser-based tools selected for their independent evaluation engines: WAVE (WebAIM), Accessibility Insights for Web (Microsoft, powered by `axe-core`), and QualWeb (Universidade de Lisboa, ACT Rules). The tools were applied to a purposive sample of 20 healthcare-related websites across five EU member states: Poland, Germany, Sweden, Finland, and Romania.

The healthcare sector was selected for two reasons. First, healthcare-related websites serve a particularly vulnerable population: people who are ill, elderly, visually impaired, or cognitively challenged are disproportionately dependent on accessible digital services. Second, public healthcare providers are already subject to WCAG 2.1 AA compliance under the Web Accessibility Directive (WAD, Directive (EU) 2016/2102), which requires public sector websites to meet accessibility requirements based on EN 301 549 (Art. 6). The EAA (Directive (EU) 2019/882), in force since 2019 with requirements applying from June 2025, complements this framework by extending comparable obligations to selected private sector products and services, such as e-commerce and consumer banking (Art. 2(2)). The five countries were selected to represent a range of national accessibility maturity, from consistently high-performing countries (Sweden, Finland) to countries with documented compliance gaps (Romania), enabling meaningful cross-country comparison.

To enable systematic cross-tool comparison on a common scale, this study proposes the Weighted Accessibility Concern Index (WACI), a metric that combines different categories of detected issues into a single weighted score. The WACI is inspired by earlier work on accessibility metrics by Pool (2023) and Nguyen (2024) but is designed specifically for browser-based tools that report errors, contrast errors, and alerts as separate categories.

Background

WCAG and EU Regulatory Framework

The Web Content Accessibility Guidelines (WCAG), developed by the W3C Web Accessibility Initiative, define the internationally accepted technical standard for accessible web content. WCAG 2.1 (2018) is the version required by EU law. The guidelines are organised around four principles (POUR): content must be Perceivable, Operable, Understandable, and Robust. Three conformance levels are defined: Level A (minimum), Level AA (legally required in the EU), and Level AAA (most stringent). The legal obligation to follow WCAG in the EU was established by Directive (EU) 2016/2102 (Web Accessibility Directive, WAD), which required all public sector websites to meet WCAG 2.1 Level AA by 2020.

Directive (EU) 2019/882 (European Accessibility Act, EAA), in force since 2019 with accessibility requirements applying from 28 June 2025 (Art. 31(2)), complements the WAD by extending comparable requirements to selected private sector products and services, including e-commerce platforms and consumer banking services (Art. 2(2)).

A detailed analysis of the legal framework and its implications for public sector compliance is addressed in a companion study focused on government websites (Maciejewska, 2026b).

Limitations of Automated Evaluation

Automated tools can detect approximately 30-40% of all WCAG violations, the remaining issues require human judgment (Gamalielsson, 2025; Pool, 2023). For example, a tool can verify that an image has an alt attribute but cannot determine whether the description is meaningful. Despite this limitation, automated evaluation is the standard approach in large-scale accessibility research and provides a reliable baseline for comparison across websites. A further limitation is that different tools use different rule sets, meaning they may detect different subsets of the same violations, or map the same violation to different WCAG success criteria (Pool, 2023).

Related Work

Research on automated web accessibility evaluation has grown significantly in recent years, particularly following the introduction of Directive (EU) 2016/2102 (Web Accessibility Directive). Studies comparing tools consistently find substantial inter-tool disagreement. Gamalielsson (2025) found that automated tools cover approximately 17% of WCAG criteria and differ substantially in what they detect, due to inconsistencies in WCAG documentation and tool implementations. Pool (2023) compared nine tools across 121 healthcare web pages and found that every tool detected issues missed by all others. In some tool pairs, combining results increased issue

detection by more than 700%. The study also proposed a pairwise divergence measure that informs the analysis in the present study. Nguyen (2024) proposed a weighted Coverage Error Ratio for tool comparison. Although developed in a US context as an honours thesis, the proposed approach is methodologically relevant to the present study.

Cross-country healthcare accessibility research provides direct context for this study. Ara and Sik-Lanyi (2022) confirmed substantial inter-tool differences across European and Asian government websites during the COVID-19 pandemic, finding that tool choice significantly affected reported compliance levels. Ara and Sik-Lanyi (2025) extended this work using automated evaluation across a broader sample, finding persistent inter-tool inconsistencies and concluding that no single automated tool provides a complete picture of accessibility compliance. Sik-Lanyi and Orbán-Mihálykó (2019) established an early precedent for cross-country healthcare accessibility comparison in the EU. Kumar et al. (2021) provide a cross-country WCAG compliance framework for government websites, though predating the EAA and not addressing inter-tool consistency.

The present study addresses gaps not covered by the above: it focuses exclusively on the healthcare sector across five EU member states in the post-EAA regulatory context, applies three browser-based tools with fully independent engines (WAVE, axe-core via Accessibility Insights, and QualWeb ACT Rules), and proposes WACI as a cross-tool comparison metric on a common scale. Unlike prior work, it explicitly addresses the absence of standardised output formats as a methodological barrier and provides replicable extraction procedures for each tool. The combination of rank-based statistical testing (Kendall's Tau) with a weighted composite metric enables both consistency assessment and burden quantification across heterogeneous tool outputs.

Evaluation Tools

The selection of evaluation tools for this study followed a systematic process. A preliminary review of nine browser-based automated accessibility evaluation tools was conducted to assess their suitability for cross-tool comparison using the WACI metric proposed in this study. Tools were evaluated against three criteria: (1) free availability without usage limitations, (2) independent underlying engine or rule set, and (3) output structure compatible with WACI (distinct E and C categories). Table 1 summarises the results of this preliminary tool review.

Table 1. Preliminary review of browser-based accessibility evaluation tools considered for this study.

Tool	Developer	Engine	E	C	A	Free	Selected	Reason for exclusion
WAVE	WebAIM	WebAIM own	Yes	Yes	Yes	Yes	Yes WACI	-
Accessibility Insights	Microsoft	axe-core	Yes	Yes	Partially	Yes	Yes WACI	Needs Review used as A(i,t), not fully equivalent to WAVE Alerts
QualWeb	Univ. Lisboa	ACT Rules + WCAG Techniques	Yes	Partially	Yes	Yes	Yes WACI	No native contrast category; C extracted via keyword filtering and manual verification (WCAG AA only)
Axe DevTools (free)	Deque Systems	axe-core	Yes	Yes	-	Yes	No	Same engine as Accessibility Insights, no Alerts in free tier
IBM Equal Access	IBM	IBM own	Yes	Yes	Yes	Yes	No	Multi-step workflow incompatible with rapid browser-based method

Lighthouse	Google	Chrome own	-	-	-	Yes	No	Reports composite score (0–100) only, no E/C/A categories
Silktide	Silktide	own	Yes	Yes	-	Partially limited	No	No Alerts equivalent, registration required for full results
LERA	accessibility.org	axe-core	Yes	Yes	-	Yes	No	Same engine as Accessibility Insights, no Alerts equivalent
BrowserStack	BrowserStack	axe-core (partial)	Yes	Yes	Partially	No paid	No	Paid tool, same axe-core engine family

Note: Yes = category available and mappable, Partially = partially available or not fully equivalent, '-' = not available. Engine independence assessed based on published tool documentation and prior literature (Gamalielsson, 2025; Pool, 2023).

Three tools were selected: WAVE (WebAIM engine), Accessibility Insights (axe-core), and QualWeb (ACT Rules; Gamalielsson, 2025), representing three independent engines. No set of three freely available tools provides fully equivalent E/C/A output, motivating the extraction procedures in Section 4.4.

WAVE

WAVE (Web Accessibility Evaluation Tool) is a browser extension by WebAIM that reports errors (definitive WCAG violations), contrast errors, and alerts (potential issues) using its own proprietary engine, independent of axe-core. It is the only tool in this study that natively provides all three WACI components (E, C, A) without extraction procedures.

Accessibility Insights for Web

Accessibility Insights for Web (Microsoft, axe-core 4.11.4) reports Failed instances with colour-contrast as a distinct subcategory, enabling direct E and C extraction. Needs Review items are included as A(i,t) in the WACI calculation. These include colour-contrast issues and other unverified findings such as duplicate-id-aria. Selected for its clear Failed/Needs Review separation and unlimited free use.

QualWeb

QualWeb (Universidade de Lisboa) implements W3C ACT Rules and WCAG 2.1 Techniques using an independent rule set. According to Gamalielsson (2025), it is the only fully free and open-source tool listed by W3C WAI. Both rule types were enabled. Failed and Warning were mapped to E and A, respectively; contrast extraction is described in Section 4.4.

Weighted Accessibility Concern Index (WACI)

The tool review established that no set of three freely available tools provides fully equivalent E/C/A output. WACI is defined as a tool-agnostic metric mapping each tool's output to a common three-component structure through replicable extraction procedures.

Let i represent a website and t a tool. For each pair (i, t) , let $\mathbf{s}(i, t) = (E(i, t), C(i, t), A(i, t))$ be the issue vector and $\mathbf{w} = (3, 2, 1)$ the weight vector. WACI is defined as:

$$\text{WACI}(i, t) = \mathbf{w} \cdot \mathbf{s}(i, t) = 3 \cdot E(i, t) + 2 \cdot C(i, t) + 1 \cdot A(i, t)$$

Weights reflect the hierarchical severity structure of WCAG conformance. Errors ($w_1 = 3$) represent definitive WCAG violations and therefore receive the highest weight. Contrast errors ($w_2 = 2$) receive an intermediate weight for two reasons. First, contrast failures are the most common accessibility issue identified in large-scale web audits, affecting 81% of homepages (WebAIM, 2025). Second, they can create significant barriers for users with low vision, a group specifically protected by EU accessibility legislation. Alerts ($w_3 = 1$) represent unverified issues requiring manual review, they may or may not constitute WCAG failures (Pool, 2023), and the lowest weight reflects this epistemic uncertainty rather than low practical importance. The 3:2:1 ratio was chosen to

preserve ordinal distinctions between categories while remaining simple enough to be replicable across tools with heterogeneous output formats. Sensitivity analysis of weight choice is a direction for future work.

Where a tool does not natively report a category, it is extracted through a replicable procedure. For QualWeb, contrast-related issues were identified by keyword filtering of the evaluation report (search term: ‘contrast’). Contrast-related failed items contribute to C(i,t), and contrast-related Warning items are excluded from A(i,t). For Accessibility Insights, colour-contrast violations are reported as a distinct subcategory of failed instances and extracted directly. Table 2 summarises the category mapping for each tool.

Table 2. Category mapping to WACI components for each tool.

Tool	E(i,t) — Errors (×3)	C(i,t) — Contrast (×2)	A(i,t) — Alerts (×1)
WAVE	Errors (native)	Contrast Errors (native)	Alerts (native)
Accessibility Insights	Failed minus colour-contrast	colour-contrast (native subcategory of Failed)	Needs Review (all items; includes colour-contrast and other unverified issues)
QualWeb	Failed minus contrast-filtered Failed	contrast-filtered Failed (keyword search: ‘contrast’)	Warning minus contrast-filtered Warning

Note: The absence of a standardised output format across tools is itself a finding of this study, consistent with Pool (2023) and Gamalielsson (2025). The replicable extraction procedures described above enable application of a single WACI formula across all three tools despite their different native classification schemes.

Methodology

Research Design

This study uses a quantitative, comparative research design based on automated accessibility auditing. The homepage of each website was evaluated independently with all three tools. The design has two analytical dimensions: (1) tool-level analysis, examining whether the three tools produce consistent results for the same pages, and (2) country-level analysis, examining whether accessibility levels differ across the five EU member states. The WACI metric enables comparison across both dimensions.

Country Selection

Five EU member states were selected based on DESI 2024 digital maturity: Sweden and Finland as top performers, Germany and Poland as mid-range performers, and Romania as a lower-performing country. All five countries are subject to Directive (EU) 2016/2102 (Web Accessibility Directive) and Directive (EU) 2019/882 (European Accessibility Act), requiring WCAG 2.1 AA compliance for public sector digital services.

Website Sample

The sample consists of 20 websites - four per country, covering four types of healthcare organizations: private clinic networks, public hospitals, national health portals, and health foundations. This typology was selected to capture variation across the public-private spectrum and different organizational missions. Only publicly accessible homepages were included. Table 3 lists the specific websites.

Table 3. Website sample by country and organization type.

Type	PL	DE	SE	FI	RO
Private clinic network	medicover.pl	medicover.de	medicover.com	terveystalo.com/fi	medicover.ro
Public hospital	szpitalczerniakowski.pl	charite.de	karolinska.se	hus.fi	suub.ro
National health portal	pacjent.gov.pl	gesundheit.de	1177.se	terveyskyla.fi	anp.gov.ro

Health foundation	ligawalkizrakiem.pl/	deutsche-herzstiftung.de	cancerfonden.se	syopasaatio.fi	fundatiarenasterea.ro
-------------------	----------------------	--------------------------	-----------------	----------------	-----------------------

Note: Medicover operates in PL, DE, and RO under separate national domains, enabling within-organization cross-country comparison. terveystalo.com/fi is the Finnish-language version of the Finnish company's .com domain. All sites were audited at homepage level only.

Data Collection Procedure

Data was collected in May 2026 using Google Chrome (version 136). For each website, all three tools were applied to the fully loaded homepage in a single session. Cookie banners were dismissed before scanning; pages were scanned without interaction with dynamic elements. For WAVE, the browser extension was activated on the loaded page, and results were read from the summary panel (Errors, Contrast Errors, Alerts). For Accessibility Insights for Web, the FastPass automated check was run and results were read from the Automated checks panel: colour-contrast violations were identified as a distinct subcategory within Failed instances, and the Needs Review count was recorded and included as A(i,t) in WACI calculations. For QualWeb, the browser extension was activated on the loaded page with both ACT Rules and WCAG Techniques enabled (default configuration), and results were read from the summary (Failed, Warning). Results were recorded in a structured spreadsheet with the following fields per tool per website: country, organisation type, URL, number of errors (E), number of contrast errors (C), number of alerts/warnings (A), WACI score.

Accessibility Scoring Method

Accessibility scores were calculated using WACI as defined in Section 4.4: $WACI(i,t) = 3 \cdot E(i,t) + 2 \cdot C(i,t) + 1 \cdot A(i,t)$. The metric was applied to all three tools using the extraction procedures described in Table 2. WACI reflects detected accessibility burden rather than full WCAG compliance.

Data Analysis

Results are analyzed at three levels: tool-level WACI scores, cross-tool divergence (Pool, 2023), and country-level means. Inter-tool consistency is assessed via Kendall's Tau (SAS PROC CORR; $\tau > 0.7$ strong, 0.4–0.7 moderate, < 0.4 weak).

Results

WACI Scores by Site and Tool

Table 4 presents WACI scores for all 20 evaluated websites across three tools. WACI scores varied substantially both across sites and across tools applied to the same site. For WAVE, scores ranged from 10 (1177.se, Sweden) to 389 (gesundheit.de, Germany), indicating a 39-fold difference in detected accessibility burden. Accessibility Insights scores ranged from 3 (karolinska.se) to 151 (medicover.pl), and QualWeb scores from 7 (anp.gov.ro) to 287 (terveystalo.fi).

Table 4. WACI scores and raw component values per website and tool (n = 20).

Website	C C	Type	WAVE				Accessibility Insights				QualWeb			
			E	C	A	WACI	E	C	A	WACI	E	C	A	WACI
medicover.pl	P L	private	9	27	24	105	2 9	2 6	12	151	2 8	3 3	1 2	162
szpitalczerniakowski.pl	P L	public	5	1	19	36	2	0	10	16	9	3	9	42
pacjent.gov.pl	P L	national-info	4	4	6	26	2	0	45	51	6	0	1 3	31
ligawalkizrakiem.pl	P L	foundation	1 3	4	15	62	8	1 2	0	48	1 2	7	1 1	61

medicover.de	D E	private	1	0	17	20	0	0	25	25	3	0	1 1	20
charite.de	D E	public	3	56	61	182	2	2 5	8	64	6	3 1	1 1	91
gesundheit.de	D E	national- info	8	13 3	99	389	3	1	11 3	124	8	3 3	1 2	102
herzstiftung.de	D E	foundation	5	0	91	106	1	0	5	8	7	0	1 2	33
medicover.com	S E	private	1 6	19	10	96	1 2	6	25	73	5	9	1 3	46
karolinska.se	S E	public	3	1	6	17	0	0	3	3	4	1	1 2	26
1177.se	S E	national- info	0	2	6	10	1	0	45	48	2	0	1 3	19
cancerfonden.se	S E	foundation	1 0	8	23	69	1 4	1	16	60	1 3	5	1 4	63
terveystalo.fi	FI	private	0	0	82	82	0	0	15	15	8 9	6	8	287
hus.fi	FI	public	2	6	7	25	1	0	43	46	6	0	1 2	30
terveyskyla.fi	FI	national- info	1 4	0	14	56	0	0	21	21	1 2	0	1 2	48
syopasaatio.fi	FI	foundation	5	9	12	45	0	7	26	40	9	7	1 0	51
medicover.ro	R O	private	2 7	12	18	123	2 8	1 4	23	135	2 2	2 7	1 2	132
suub.ro	R O	public	1 0	47	22	146	8	2 8	36	116	4	3 0	1 4	86
anp.gov.ro	R O	national- info	3 1	17	14 2	269	2 9	6	42	141	2	0	1	7
fundatiarenasterea .ro	R O	foundation	4 1	55	43	276	3 0	1 8	0	126	4 0	5 2	9	233

Note: CC = country code. WAVE: E = Errors, C = Contrast Errors, A = Alerts (all native). Accessibility Insights: E = Failed minus colour-contrast, C = colour-contrast (native subcategory), A = Needs Review, WACI = $3 \cdot E + 2 \cdot C + 1 \cdot A$. QualWeb: E = Failed minus contrast-filtered items, C = contrast-related Failed (keyword: 'contrast', WCAG AA level), A = Warnings, WACI = $3 \cdot E + 2 \cdot C + 1 \cdot A$.

Tool Consistency: Kendall's Tau

Rank correlations between tool pairs were calculated using Kendall's Tau to assess whether the three tools produce consistent site rankings despite differences in absolute WACI values. Table 5 presents the results.

Table 5. Kendall's Tau rank correlations between WACI scores across tool pairs (n = 20 websites).

Tool pair	Kendall's τ	p-value	Interpretation
WAVE vs Accessibility Insights	0.449	0.006	Moderate, significant
WAVE vs QualWeb	0.516	0.001	Moderate, significant
Accessibility Insights vs QualWeb	0.343	0.035	Weak-moderate, significant

Note: All correlations statistically significant at $\alpha = 0.05$. τ = Kendall's rank correlation coefficient.

All three pairwise correlations were statistically significant ($p < 0.05$). The highest agreement was observed between WAVE and QualWeb ($\tau = 0.516$, $p = 0.001$), followed by WAVE and Accessibility Insights ($\tau = 0.449$,

$p = 0.006$), and Accessibility Insights and QualWeb ($\tau = 0.343$, $p = 0.035$). These results indicate moderate but consistent agreement in site ranking across tools, suggesting that different tools tend to identify similar relative accessibility patterns despite differences in absolute WACI values. 6.3 Cross-Country Comparison. Table 6 presents mean WACI scores by country for each tool.

Table 6. Mean WACI scores by country across three tools (n = 4 websites per country).

Country	WAVE mean	WAVE rank	AI mean	AI rank	QualWeb mean	QualWeb rank
Sweden	48.0	1	46.0	2	38.5	1
Finland	52.0	2	30.5	1	104.0	4
Poland	57.2	3	66.5	4	74.0	3
Germany	174.3	4	55.2	3	61.5	2
Romania	203.5	5	129.5	5	114.5	5

Note: n = 4 websites per country (one private, one public hospital, one national health portal, one health foundation). Rank = ranking by mean WACI (1 = lowest burden, 5 = highest). Rankings calculated independently per tool.

According to WAVE, Sweden recorded the lowest mean WACI (48.0), followed by Finland (52.0), Poland (57.2), Germany (174.3), and Romania (203.5). Despite differences between tools, Sweden consistently ranked among the best-performing countries, while Romania recorded the highest mean WACI across all three tools. However, tool choice affected the ranking of mid-range countries. Finland ranked first by Accessibility Insights (30.5) but fourth by QualWeb (104.0), driven by the *terveystalo.fi* anomaly discussed in Section 6.4. Germany ranked fourth by WAVE but second by QualWeb (61.5). Germany also showed substantial tool divergence, with a mean WACI of 174.3 in WAVE compared with 55.2 in Accessibility Insights. This difference was driven primarily by *gesundheit.de* (WACI_WAVE = 389, WACI_AI = 124), where WAVE detected substantially more alerts. These ranking reversals indicate that absolute WACI values and country-level comparisons are tool-dependent.

Cross-Tool Divergence

The five most divergent sites by absolute WAVE–QualWeb WACI difference were: *gesundheit.de* (287), *anp.gov.ro* (262), *terveystalo.fi* (205), *charite.de* (91), and *herzstiftung.de* (73). The five most consistent sites were: *medicover.de* (difference = 0), *ligawalkizrakiem.pl* (1), *pacjent.gov.pl* (5), *hus.fi* (5), and *cancerfonden.se* (6). The case of *terveystalo.fi* illustrates cross-tool divergence: WAVE reported WACI = 82, Accessibility Insights WACI = 15, and QualWeb WACI = 287. WAVE detected 82 alerts (predominantly structural), Accessibility Insights detected 15 issues, while QualWeb flagged 95 errors under ACT Rules, reflecting focus-management failures not captured by axe-core-based tools. This divergence demonstrates that different rule sets surface genuinely different accessibility concerns.

Discussion

The results provide three interconnected findings. First, all three tool pairs produced statistically significant rank correlations ($\tau = 0.343$ – 0.516), indicating that despite different underlying engines and rule sets, the tools agree on which sites have relatively higher versus lower accessibility burden. This supports the validity of WACI as a cross-tool ranking metric: while absolute values are tool-dependent, relative site ordering is moderately consistent.

Second, absolute WACI values differed substantially across tools for the same website. WAVE scores ranged from 10 to 389, while the largest tool-specific divergence was observed for *terveystalo.fi*, where QualWeb reported a WACI of 287 compared with 15 in Accessibility Insights. These differences indicate that automated tools should not be used interchangeably when comparing absolute accessibility burden and highlight the importance of methodological transparency in accessibility audits. This is consistent with Pool (2023) and Gamalielsson (2025), who document substantial inter-tool disagreement. The proposed WACI metric enables comparison across heterogeneous tool outputs, and Kendall's Tau enables rank-based assessment of inter-tool consistency. The *terveystalo.fi* case deserves particular attention as the most striking within-dataset divergence. QualWeb detected 95 failed items under ACT Rules, primarily focusing on focus management and keyboard navigation issues tested through scripted interaction sequences that axe-core does not cover. Accessibility Insights, which uses the axe-core engine, reported only 15 WACI for the same site. For a Finnish healthcare provider audited under EU accessibility requirements, this tool choice alone would change the compliance

assessment by an order of magnitude. This finding has direct implications for auditors and regulators: tool selection is not a neutral methodological choice, it determines which WCAG success criteria are tested and therefore which failures are visible.

Third, the country-level pattern: Romania highest, Sweden and Finland lowest, is consistent with DESI 2024 digital maturity rankings. Sweden and Finland, with active monitoring bodies and published accessibility reports, show systematically lower WACI across all tools. Romania showed the highest WACI across all tools, suggesting potential differences in enforcement maturity consistent with its lower DESI ranking. This pattern is only partially consistent with DESI 2024 rankings: Germany, which ranks highly on digital maturity, showed the second-highest mean WACI (174.3, WAVE), driven primarily by *gesundheit.de* (WACI = 389). This suggests that national digital maturity does not automatically translate into lower automated accessibility burden - sector-specific factors and enforcement focus may matter more than aggregate digital development. The relationship between enforcement maturity and accessibility outcomes will be examined in a companion study (Maciejewska, 2026b).

Limitations

Several limitations should be noted. Automated tools detect approximately 30–40% of WCAG violations (Gamalielsson, 2025). WACI captures only the burden these tools can observe, not a complete WCAG compliance assessment.

The sample of 20 websites, while purposive, is not statistically representative of the healthcare sector in each country. Dynamic content and third-party widgets (including accessibility overlay tools observed on some evaluated sites) may affect automated results. One website (*suub.ro*) was used in place of the originally selected hospital site after repeated QualWeb evaluation failures. The replacement represents the same organisation type and country. The category labelled "national health portal" includes both government-operated portals and nationally recognised public health information services. Websites were classified according to their functional role as publicly accessible sources of health information rather than their ownership structure.

Conclusions

This study compared three browser-based automated accessibility evaluation tools: WAVE, Accessibility Insights for Web, and QualWeb, applied to 20 healthcare-related websites across five EU member states. The proposed Weighted Accessibility Concern Index ($WACI = 3 \cdot E + 2 \cdot C + 1 \cdot A$) enabled systematic comparison across tools with heterogeneous output formats through replicable extraction procedures. Results showed moderate but statistically significant rank agreement between all tool pairs (Kendall's $\tau = 0.343$ – 0.516 , all $p < 0.05$). Browser-based tools generally agree on relative accessibility burden while producing substantially different absolute WACI scores for the same sites. Tools should be used comparatively rather than interchangeably, with explicit documentation of the tool and extraction procedure applied.

Cross-country analysis revealed that Sweden and Finland consistently achieved the lowest mean WACI, while Romania recorded the highest across all tools. Germany showed unexpectedly high accessibility burden relative to its DESI 2024 ranking, indicating that national digital maturity alone does not guarantee lower automated accessibility burden.

Future work will extend the comparison to CLI and API-based tools and investigate accessibility-related procurement practices. The findings carry practical implications for three groups. For accessibility auditors: tool selection is not methodologically neutral, results should always document the tool used, its version, and the extraction procedure applied. WACI enables cross-tool comparison where multiple tools are used. For regulators and supervisory bodies enforcing the EAA: automated tool results alone are insufficient for compliance determination - standardised methodology and tool-agnostic metrics are needed for consistent cross-country monitoring. For tool developers: the absence of standardised output formats is a barrier to research reproducibility and cross-tool comparison - harmonisation of E/C/A categories would substantially advance the field.

References

- Ara, J. and Sik-Lanyi, C. (2022) 'Investigation of COVID-19 vaccine information websites across Europe and Asia using automated accessibility protocols.' *International Journal of Environmental Research and Public Health*, 19(5), 2867. <https://doi.org/10.3390/ijerph19052867>
- Ara, J. and Sik-Lanyi, C. (2025) 'Automated evaluation of accessibility issues of webpage content: tool and evaluation.' *Scientific Reports*, 15, 9516. <https://doi.org/10.1038/s41598-025-92192-5>

- European Commission. (2024) *Digital Economy and Society Index (DESI) 2024*. Publications Office of the European Union. <https://digital-strategy.ec.europa.eu/en/policies/desi>
- European Parliament and Council of the European Union. (2016) *Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies*. Official Journal of the European Union. <https://eur-lex.europa.eu/eli/dir/2016/2102/oj>
- European Parliament and Council of the European Union. (2019) *Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services (European Accessibility Act)*. Official Journal of the European Union. <https://eur-lex.europa.eu/eli/dir/2019/882/oj>
- Gamalielsson, J. (2025) 'Coverage of web accessibility guidelines provided by automated checking tools.' *Universal Access in the Information Society*, 24(4), 3615–3637. <https://doi.org/10.1007/s10209-025-01263-x>
- Kumar, D., Biswas, P. and Venkatesh, M. (2021) 'Government website accessibility: a cross-country analysis of G7 and BRICS countries.' *Universal Access in the Information Society*. <https://doi.org/10.1007/s10209-021-00804-4>
- Maciejewska, K. (2026b) 'Accessibility Declarations vs. Automated Audit Results: A Cross-Country Study of Government Websites in Six EU Member States.' In *Proceedings of the 47th International Business Information Management Association (IBIMA)*, Madrid, Spain. ISSN 2767-9640. [in press]
- Nguyen, T. (2024) 'Evaluating automated accessibility checker tools.' *WWU Honors College Senior Projects*, 805. https://cedar.wvu.edu/wwu_honors/805
- Pool, J. R. (2023) 'Accessibility metatesting: comparing nine testing tools.' In *Proceedings of the 20th International Web for All Conference (W4A '23)*. <https://doi.org/10.1145/3587281.3587282>
- Sik-Lanyi, C. and Orbán-Mihálykó, É. (2019) 'Accessibility testing of European health-related websites.' *Arabian Journal for Science and Engineering*, 44(11), 9171–9190. <https://doi.org/10.1007/s13369-019-04017-z>
- W3C Web Accessibility Initiative. (2018) *Web Content Accessibility Guidelines (WCAG) 2.1*. World Wide Web Consortium. <https://www.w3.org/TR/WCAG21/>
- WebAIM. (2025) *The WebAIM Million: The 2025 report on the accessibility of the top 1,000,000 home pages*. <https://webaim.org/projects/million/2025>