Sensitivity and specificity of EtG in hair as a marker of chronic excessive drinking. Pooled analysis of raw data and meta-analysis of diagnostic accuracy studies

This is the author's manuscript

Original Citation:
Sensitivity and specificity of EtG in hair as a marker of chronic excessive drinking. Pooled analysis of raw data and meta-analysis of diagnostic accuracy studies / Rafael Boscolo-Berto; Donata Favretto; Giovanni Cecchetto; Marco Vincenti; Robert Kronstrand; Santo Davide Ferrara; Guido Viel. - In: THERAPEUTIC DRUG MONITORING. - ISSN 0163-4356. - STAMPA. - 36:5(2014), pp. 560-575.

Availability:
This version is available http://hdl.handle.net/2318/155920 since 2015-12-29T09:37:46Z

Published version:
DOI:10.1097/FTD.0000000000000063

Terms of use:
Open Access
Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)
This is an author version of the contribution published on:

Therapeutic Drug Monitoring, 36, 2014, DOI: 10.1097/FTD.0000000000000063.

Volume 36, LWW Journal, 2014, 560-575

The definitive version is available at:

Abstract

Background: To assess the debated diagnostic performance of ethyl glucuronide in the 3-cm proximal scalp hair fraction (HEtG) as a marker of chronic excessive drinking.
Methods: In July 2012/May 2013, after a systematic search through the MEDLINE, OVID/EMBASE, WEB OF SCIENCE, and SCOPUS databases, 8 studies were included in the pooled analysis that report raw single data on HEtG concentration and self-reported daily alcohol intake (SDAI). A receiver operating characteristic curve analysis and a Spearman rank-order correlation test were used. A meta-analysis was performed following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses and Cochrane recommendations, comprising quality and bias assessments. Results: The pooled analysis showed that 30 pg/mg could be a useful cutoff value for HEtG to detect an SDAI >60 g/d and demonstrated a parabolic direct correlation between HEtG and SDAI data [rho 0.79; 95% confidence interval (CI), 0.69–0.87; P < 0.001]. The meta-analysis found an overall HEtG sensitivity of 0.96 (95% CI, 0.72–1.00) and a specificity of 0.99 (95% CI, 0.92–1.00); a nomogram to predict the posttest probability of exhibiting the targeted condition in the general population was built. Significant variability among the included studies was detected, which was mainly explained by true heterogeneity in the presence of publication bias.
Conclusions: With the available data, we conclude that HEtG is a promising marker for identifying chronic excessive drinking. Nonetheless, larger and well-designed population studies are required to draw any definitive conclusions on the significance and appropriateness of its application in the forensic setting.

Key Words
alcohol abuse, ethyl glucuronide, EtG, hair, meta-analysis

Introduction

Efficient and reliable biomarkers of alcohol consumption are necessary in both clinical and research arenas for differentiating moderate/social from heavy/problematic drinking, evaluating treatment programs and clinical trials, and determining recent drinking episodes in periods of required abstinence (ie, organ transplantation, withdrawal of driving license, workplace risk prevention, pregnancy, etc.).[1–3] Especially in the forensic context, where the self-reported daily alcohol intake (SDAI) is generally unreliable, as the individuals are motivated to deny or minimize the magnitude of their drinking behavior to avoid legal implications, sensitive and specific biomarkers of chronic excessive drinking are needed.[4–8]
In the last decade, special focus has been addressed to the quantification of ethanol metabolites in keratin matrices [ie, ethyl glucuronide (EtG) and fatty acid ethyl esters], resulting in a high number of articles published on this topic in journals of legal medicine and forensic toxicology, particularly by European research groups.[9–23]

In particular, it has been proposed that EtG, a phase 2 non-oxidative minor metabolite of ethanol produced by the endoplasmic reticulum Uridine 5'-diphospho-glucuronosyltransferase, could provide a direct measure of alcohol exposure over an extended detection time window, overcoming the main limitations related to indirect alcohol biomarkers. [2,17,22,24–26] Although the mechanism of EtG incorporation into the hair shaft has been only partially elucidated, recent experimental data gathered on healthy volunteers have shown the predominant role of blood incorporation with respect to sweat and sebum incorporation. [5,6,23,27] Because of its acidic and hydrophilic nature, EtG is unlikely to be bound to melanin, and thus, its affinity for keratinocytes should not be influenced by natural hair color. It has been shown that aggressive chemical treatments (ie, hair bleaching, perming, and coloring) can decrease hair EtG (HETG) concentration, depending mainly on the chemical composition of the cosmetics used and the frequency of their application. [28–31]

In both clinical and forensic literature, several cutoff values (4, 7, 23, 25, 27, 30, and 50 pg/mg) have been proposed for HETG concentration to indicate chronic excessive drinking, although the majority of these studies examined only selected groups of individuals.[11,13,16,20,27,32–35]

In 2009, the Society of Hair Testing (SoHT) approved a consensus document (revised in 2011) reporting that HETG concentration in the 0- to 3-cm up to 0- to 6-cm proximal scalp hair segment can be used for diagnosing a chronic excessive drinking behavior, defined as an average consumption of at least 60 g of pure ethanol per day over several months and that the best HETG cutoff to be used for that purpose is 30 pg/mg. [36–38]

This document raised some concerns, as several authors pointed out that considerable interindividual variability is expected for HETG and that multicenter studies (promoted by international scientific societies or by independent multicentric cooperative centers) are required for characterizing the distribution of HETG concentration in nondrinkers and social drinkers. [5,19,39–41]

The present article aims to the systematic review of the literature with the following purposes.

1. Aggregate all published raw data on individual HETG concentrations in subjects with known SDAI to perform pooled and meta-analyses and identify the cutoff value that provides the best diagnostic accuracy in terms of sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and likelihood ratios for detecting an alcohol intake higher than 60 g/d. This value has been widely proposed in both clinical and forensic literature as the threshold to distinguish a social from a heavy drinking behavior. [6,25,36,37,42]
2. Investigate the correlation between HEtG and SDAI data, calculating the degree of variance of the HEtG concentration (dependent variable) explained by the SDAI (independent variable).
3. Verify the possible existence of biases affecting the cohorts used in the meta-analyzed studies and predict the diagnostic performance of the HEtG marker out of the experimental setting (i.e., in a real clinical or forensic setting).

**Materials and Methods**

To pursue the above-mentioned aims, the following multistep analytical strategy was used.

**Systematic Search Strategy**

In July 2012, 2 authors (R.B.B. and G.V.) performed the Web-based systematic search of the literature through the databases MEDLINE/PUBMED, OVID/EMBASE, WEB OF SCIENCE, and SCOPUS to identify relevant articles on the diagnostic performance of the HEtG marker.43 The search terms were kept intentionally broad, without adoption of temporal limits or language restrictions, to be as sensitive as possible and avoid misleading results.[44] The MEDLINE/PUBMED, OVID/EMBASE, and SCOPUS searches were performed by an inquiry strategy, including “free-text” protocols, including the terms “EtG,” “ethyl-beta-D-glucopyranosiduronic acid,” “ethyl beta-D-6-glucosiduronic acid,” “ethyl glucuronide,” “ethylglucuronide,” or “ethyl-glucuronide” and “hair.” As a consequence of the Web interface limitations, the search fields for WEB OF SCIENCE were restricted to “topic” and “title.” A comprehensive database of the retrieved articles was built and manually checked for removing any duplicates. To further search potentially relevant articles, the reference lists of included studies were screened so as the manuscripts published after the systematic search until the week before submission. A final check to update the systematic search was repeated just before the article submission (May 2013) to include any new contribution on this issue.

**Article Selection**

**Inclusion Criteria for Pooled and Meta-Analyses**

1. Provide single-case data on EtG concentration in human scalp hair measured with a validated chromatographic mass spectrometric method and data on the sensitivity and/or specificity, true- and/or false-positive rates, true- and/or false-negative rates, and PPV and/or NPV of the HEtG
marker for the detection of chronic alcohol ingestion of more than 60 g/d, that is, chronic excessive drinking.

2. Report a limit of quantification (LOQ) for the analytical method $\leq 10$ pg EtG/mg hair.

3. SDAI data of the included subjects for at least 1 month before the EtG analysis.

**Exclusion Criteria for Pooled and Meta-Analyses**

Articles not fulfilling all the previous requirements, being letters to the editor or reviews, reporting formats of data not calculable/transformable by published results/figures, or unsuitable for direct processing by pooled/meta-analysis, were excluded. The study selection was independently performed on title and abstract by 3 authors (R.B.B., G.C., and G.V.), according to the above-mentioned inclusion and exclusion criteria. Studies considered relevant by at least 2 of the 3 authors were selected and subsequently examined in full text. Disagreements were adjusted by discussion, basing the consensus on the obtained full texts, whereas persisting discrepancies were solved by consulting a fourth author (S.D.F.).

**Data Extraction**

Data extraction was conducted independently by 2 authors (R.B.B. and G.C.) on articles retrieved by the systematic search and classified relevant for the pooled or the meta-analysis and separately collected data from full text manually compiling an electronic database, whereas a third author (G.V.) checked the accuracy and agreement of the extracted data to minimize subjective evaluation. The following items were collected from each study and inserted into a predefined table: first author, publication year, main aim, inclusion/exclusion criteria, number of subjects, experimental setting, subjects’ stratification, mode of data acquisition for SDAI, SDAI value for at least 1 month before test, hair length, analytical method with its limit of detection and LOQ, HEtG concentration, sensitivity, specificity, PPV, NPV, and main finding of the study. Discrepancies in the data extraction process were settled through consensus discussion, whereas persisting divergences were solved by consulting a fourth author (S.D.F.).

**Statistical Analyses**

One of the authors (R.B.B.) performed the pooled analysis and the meta-analysis according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statements and the Cochrane Collaboration recommendations.[45]
Pooled Analysis

The accuracy of the HEtG test was assessed through receiver operating characteristic (ROC) curve analysis with area under the curve (AUC) and 95% confidence intervals (CIs), using SDAI >60 g/d as the target category. The Kolmogorov–Smirnov test was used to assess whether the distributions of values did approximate a normal Gaussian curve. The degree of association between the quantitative continuous dependent variable (HEtG) and the independent variable (SDAI) was assessed through the Spearman rank-order correlation (rho), according to the nonparametric distribution of values (see above). We also calculated the regression line and the coefficient of determination (r²) to evaluate how much the variance of the dependent variable (HEtG) is explained by the independent variable (SDAI). A 2-tailed P value <0.05 was considered statistically significant.

Meta-Analysis

Statistical analyses were dependent on the completeness data for true/false positives and true/false negatives, allowing the meta-analytic computation of sensibility, specificity, PPV, and NPV. To avoid any potential model misspecification, we preliminarily checked for [chi]² probability plot of squared Mahalanobis distances to assess the bivariate normality assumption and a bivariate boxplot to evaluate the degree of interdependence of sensitivity and specificity. Meta-analysis of random effects estimates was performed to investigate the change in the overall effect size (by Cohen d standardized mean difference test) and 95% CIs for the meta-analysis, after eliminating one study at a time. Presence of heterogeneity was assessed by Cochrane-Q test, with subsequent quantification by the statistical index of heterogeneity (I²), which represents the percentage of the total variability in a set of effect sizes because of true heterogeneity (ie, between-study variability). Pooled estimates of sensitivity and specificity with corresponding 95% CIs were analyzed based on the bivariate model for diagnostic meta-analysis to obtain an overall sensitivity and specificity. [46,47] Random or fixed models were used depending on the results of the heterogeneity test (see above). To graphically present the results, the individual and summary points of sensitivity and specificity were plotted on an ROC graph, with the sensitivity (true-positive rate) of the test on the y axis and the false-negative rate on the inverted x axis. In addition, the 95% confidence region and 95% prediction region around the pooled estimates were added to illustrate the precision with which the pooled values were estimated (confidence ellipse of a mean) and to show the “between-study variation” (prediction ellipse, the likely range of values for a new study).

The likelihood ratio scattergram was elaborated as a function of sensitivities and specificities to better understand the diagnostic value of the HEtG test in predicting SDAI data. The Fagan
nomogram was used to calculate the positive and negative posterior probabilities based on previous probabilities (i.e., prevalence of the targeted condition in the investigated population). To assess the potential publication bias, the Begg–Mazumdar adjusted rank correlation analysis, the Egger regression of standardized effect estimates against their precision, the Harbord modified Egger test, and the Duval and Tweedie nonparametric analysis were performed. For all analyses, a 2-sided $P < 0.05$ was adopted to indicate statistical significance.

Sample Size and Power Analysis

A theoretical prevalence for the investigated condition (SDAI > 60 g) in the general population (http://www.euro.who.int/en/what-we-do/health-topics/disease-prevention/alcohol-use/country-work/country-profiles) ranging from 5% to 10% was hypothesized for sample size and power analysis. A 2-sided binomial test with a minimum significance level of 0.05 was used to calculate the required sample size for an hypothetical experimental study aimed at achieving an overall statistical power ranging from 0.80 to 0.95 for the detection of changes in the sensitivity of the HEtG test from 0.50 (nondiscriminating test) to 0.80 and in the specificity of the HEtG test from 0.50 to 0.98. [52–54]

Assessment of the Quality of the Studies

All the studies included in the pooled/meta-analysis underwent quality assessment, as previously suggested.[45,55] Two authors (R.B.B. and G.V.) independently rated the methodological quality of the included articles according to the 4 domains Quality Assessment of Diagnostic Accuracy Studies-2, a validated tool for the quality assessment of the diagnostic accuracy of eligible studies. [56] As indicated in the original article, the tool was modified by adding 5 tailored signaling questions to better assess the risk of bias. [56]

Results

Systematic Search

The combined search in the databases MEDLINE/PUBMED, OVID/EMBASE, WEB OF SCIENCE, and SCOPUS retrieved 578 plus 3 records, 212 of which were excluded being duplicates (Fig. 1). Of the remaining 369 records screened by title and abstract, 309 were excluded as they were judged not pertinent to the topic, whereas 60 articles were examined in full text.
Nine (2.5%) studies fully matched the selection criteria and were chosen for the pooled or meta-analysis (Table 1). Nine other pertinent studies reported data unsuitable for quantitative analyses; therefore, their findings were only summarized in Table 2. In reviewing the extracted data, there was complete agreement among the authors.

Pooled Analysis

Eight out of the 9 above-mentioned studies were included in the pooled analysis (Table 1). Complete paired data (SDAI and HEtG) were available for 70 subjects, which were included in the raw data analysis. A statistically significant difference in the SDAI (P < 0.05) and HEtG concentrations (P < 0.05) was found between the 2 considered drinking categories (SDAI > 60 g for heavy drinkers or SDAI < 60 g for social drinkers). Subjects self-reporting a daily intake <=60 g of ethanol (n = 28; mean SDAI = 25.2 g with 95% CI, 19.2–31.2 g; SD 15.4 g) exhibited a mean HEtG value of 11.2 pg/mg (95% CI, 7.7–14.7 pg/mg; SD 9.0 pg/mg), whereas heavy drinkers (n = 42; mean SDAI = 190.5 g with 95% CI, 151.1–229.9 g) showed a median HEtG concentration of 51.5 pg/mg (95% CI, 38.4–86.2 pg/mg; interquartile range 30–140 pg/mg). Direct correlation between HEtG and SDAI data with a Spearman rank correlation coefficient (rho) of 0.79 (95% CI, 0.69–0.87; P < 0.0001) was demonstrated. The regression analysis led to a coefficient of determination (r2) of 0.31 (P < 0.001 by F test for variance), explained by the following equation: HEtG = 8.1895 + 0.5234 × SDAI - 0.0003224 × SDAI^2 (Fig. 2). On this basis, both sensitivity and specificity indexes were calculated and plotted against any possible HEtG concentration. The reference parameter used for determining the diagnostic efficiency of the HEtG test was SDAI, with the “positive case” classified as a subject self-reporting to drink more than 60 g/d (Fig. 3). Furthermore, serial sensitivity, specificity, and positive and negative likelihood ratios were calculated for any possible HEtG cutoff value above the threshold of 10 pg/mg, that is, the maximum admissible LOQ for studies to be included into the present pooled analysis (Table 3). The 30 pg/mg cutoff led to the best diagnostic efficiency, with a specificity of 0.97 (95% CI, 91.6–99.4) and a sensitivity of 0.85 (95% CI, 74.6–92.2).

Meta-Analysis

Eight out of the 9 selected studies were included in the meta-analysis (Fig. 1). The preliminary check for the sensitivity analysis showed no significant changes after the systematic exclusion of each of the selected studies at a time.

The overall sensitivity of the HEtG test for detecting the target condition (SDAI > 60 g/d) was 0.96 (95% CI, 0.72–1.00) (Fig. 4). Significant variability was detected among the included studies, with 94.4% (95% CI, 91.8–97.0) of the variability explained by true heterogeneity (Cochrane-Q test
The overall specificity of the HEtG test for detecting an SDAI above 60 g/d was 0.99 (95% CI, 0.92–1.00) (Fig. 4). Significant variability was detected among the studies, with 66.1% (95% CI, 40.5–91.6) of the variability explained by true heterogeneity (Cochrane-Q test 20.6; P < 0.00). The combined sensitivity and specificity of each study generated an overall ROC curve, which offers a synthesis of the test performance and displays the trade-off between its sensitivity and specificity. Unlike the traditional ROC curve, which explores the effect of varying the decision thresholds (ie, the cutoff values) on the sensitivity and specificity of the test by considering individual data points, our graph shows aggregated data (ie, each plot represents a study). The resulting curve is positioned near the desirable upper left corner of the graph with an AUC of 0.99 (95% CI, 0.98–1.00) (Fig. 5). Despite this punctual estimation, the 95% confidence region for the aggregated sensitivity and specificity is relatively wide, extending from 0.38 to 1.00 for sensitivity and from 0.80 to 1.00 for specificity. The prediction region ranges from 0.01 to 1.00 for sensitivity and from 0.65 to 1.00 for specificity (Fig. 5).

The likelihood ratio scattergram displayed the pooled positive and negative likelihood ratio points at 28.9 (95% CI, 26.1–32.0) and 0.16 (95% CI, 0.04–0.5), respectively (Table 3), with an unconditional PPV of 0.98 (95% CI, 0.97–1.00) and unconditional NPV of 0.96 (95% CI, 0.94–0.97). To achieve the above-reported sensitivity of 0.85 (95% CI, 74.6–92.2) and specificity of 0.97 (95% CI, 91.6–99.4), both referred to the cutoff of 30 pg/mg emerging from our pooled analysis (Table 3), the positive likelihood ratio and negative likelihood ratio values were set at 28.9 and 0.16, respectively. With these settings, the summary point fell into the left upper quadrant, which means that the test is theoretically useful for the confirmation of the targeted condition (when really present) and for excluding it when negative (SDAI < 60 g/d). Nevertheless, the summary point shows a CI that crosses over all the other quadrants, making its localization merely indicative (Fig. 6A). Additionally, given the pretest probability of the targeted condition (ie, the prevalence of SDAI > 60 g/d in the investigated population), a nomogram to determine the posttest probabilities was built. The overall prevalence of SDAI above 60 g/d in the included studies was 48.3% ± 19.3% (mean and SD), being coherent with the prevalence of heavy drinkers (42%) in the available raw data (included in the pooled analysis). Hence, fixing the pretest probability at 42%, positive and negative posttest probabilities of 98% and 3%, respectively, were obtained (Fig. 6B). Because the prevalence of heavy drinkers (SDAI > 60 g) is usually lower than 10% in the general population (http://www.euro.who.int/en/what-we-do/health-topics/disease-prevention/alcohol-use/country-work/country-profiles), we calculated that a posttest probability to really face a subject with the targeted condition is equal to 80%–90% when the test is positive and 0.2%–0.5% when the test is negative.
Sample Size and Power Analysis

Considering that the mean prevalence of heavy drinkers in the general population is reasonably expected to range from 5% to 10% (http://www.euro.who.int/en/what-we-do/health-topics/disease-prevention/alcohol-use/country-work/country-profiles; last access on December 18, 2013), minimum theoretical sample sizes were calculated to allow appropriate evaluation on statistical significance for the combined HEtG sensitivity of 0.80 and specificity of 0.98 against a theoretical nondiscriminating test (ie, null hypothesis with specificity and sensitivity equal to 0.50%), by means of a 2-sided binomial test at various power values (Table 4).

In detail, with a pretest power fixed at 0.80, a minimum total sample size of 400 subjects (including 20 heavy drinkers) or 200 subjects (including 20 heavy drinkers) is required with an indicative hypothetical prevalence of the target condition within the general population equal to 5% or 10%, respectively. On the other hand, raising the pretest power at 0.95, a minimum total sample size of 700 subjects (including 35 heavy drinkers) or 350 subjects (including 35 heavy drinkers) is required with an indicative hypothetical prevalence of the target condition within the general population equal to 5% or 10%; respectively (Table 4).

Quality Assessment

The preliminary quality assessment of the meta-analyzed studies (Table 1) was conducted by means of the Quality Assessment of Diagnostic Accuracy Studies-2 tool and showed no major concerns of applicability, in terms of reference standard (100% of studies with low concern), index test (100% low concern), and patient selection (100% low concern).

Despite the fulfillment of these qualitative eligibility criteria, a subsequent more in-depth evaluation for potential bias showed a number of significant drawbacks. The patient selection showed a high risk of bias for 22.2% of the articles, the interpretation of the HEtG analysis showed a high risk of bias for 22.2% of the studies, the interpretation of the reference standard (eg, self-reported consumption, Alcohol Use Disorders Identification Test, TimeLine FollowBack, etc.) implied a high risk of bias for 77.8% of the studies, and the flow and timing of analysis showed a high risk of bias for 44.4% of the included studies.

The Begg–Mazumdar adjusted rank correlation test was applied to evaluate the publication bias (ie, tendency on average to produce results that seem significant because positive results have a better chance to be published, to be published earlier, and in journals with a higher impact) and displayed a significant asymmetry (P < 0.05 after correction for continuity) in the presence of a low statistical power (see Figure C, Supplemental Digital Content 1, http://links.lww.com/TDM/A83). There was evidence of a lack of published articles in the lower right side, which evidences missing studies with low effect and high variance (see Figure C, Supplemental Digital Content 1,
The Egger test for funnel plot asymmetry confirmed a significant deviation from the origin of the axes (bias 1.97; 95% CI, 1.25–2.69; P < 0.001) showing the presence of small studies, which overestimate the effect, or, alternatively, the absence of negative/nonsignificant small studies (see Figure C, Supplemental Digital Content 1, http://links.lww.com/TDM/A83). This point was further investigated by performing the Duval–Tweedie nonparametric test (see Figure C, Supplemental Digital Content 1, http://links.lww.com/TDM/A83). The theoretical absence of at least 4 studies, all 1-sided and located in the regions of low/no statistical significance (P > 0.10) with a negative effect and a moderate standard error, was identified.

**Discussion**

The aims of the present pooled and meta-analyses were (1) to investigate the correlation between EtG concentration in scalp hair and daily ingested amount of ethanol, (2) to identify the cutoffs that maximize the diagnostic accuracy of the marker, and (3) to verify the possible existence of biases affecting the cohorts of any of the included studies.

A direct correlation between HETG concentrations and the SDAI was demonstrated with a Spearman rank correlation coefficient of 0.79 (95% CI, 0.69–0.87; P < 0.0001) (Fig. 2). The positive Spearman correlation index indicates that the HETG concentration increases with the SDAI. This index increases in magnitude (from 0 to 1) as the SDAI and the HETG become closer to each other, being perfect monotone functions with a value of 1. Our results are in partial agreement with the previously published literature data. Politi et al 11 using a second-grade polynomial function reported a coefficient of correlation (r²) of 0.86; Appenzeller et al 27 found a Pearson product moment correlation (rp) of 0.5357, whereas more recently, Lees et al 19 reported a Spearman rank coefficient of 0.42. Interestingly, our regression analysis performed on the pooled SDAI and HEtG data led to a coefficient of determination (r²) of 0.31 (P < 0.001 by F test for variance), demonstrating that only 31% of the variance of the HEtG concentration (dependent variable) is explained by the SDAI (independent variable). These data suggest that other currently underestimated variables, such as those related to the ethanol metabolism (age, enzyme activity, drinking pattern, etc.) and/or to the incorporation of the EtG molecule into the hair shaft (ie, sweating, production of sebum, cosmetic treatments, hygienic habits, etc.), may significantly influence the HEtG concentration.11,13 Moreover, the best mathematical equation describing the relationship between SDAI and HETG data is a parabolic function, not a straight line. Two alternative hypotheses might be formulated for interpreting this specific issue (ie, plateau of the graph in Fig. 2), namely, the saturation of the Uridine 5′-diphospho-glucuronosyltransferase enzyme at high blood alcohol concentrations or the equilibrium between incorporation and washout effects, which at high HEtG concentrations might equal each other (Fig. 2).
As recently described by our research group, alcohol abusers, social drinkers, and teetotalers can exhibit partially overlapping HEtG concentrations. Therefore, the proper choice of the cutoff value effectively discriminating moderate from nonmoderate (SDAI > 60 g) alcohol users strictly depends on the degree of sensitivity and specificity needed for the decision.

Several cutoff values, mainly derived from the case–control studies involving a limited number of selected individuals, have been proposed in the literature. In 2009, Morini et al. investigating 75 subjects with an SDAI of more than 60 g in the last 3 months and 23 social drinkers (SDAI < 60 g) found that the 27 pg/mg threshold provided the best compromise between sensitivity (0.92) and specificity (0.96) for detecting the target condition (ie, SDAI > 60 g). Appenzeller et al., using the correlation equation between SDAI and HEtG data calculated for 15 withdrawal treatment patients, found that 23 pg/mg was the best threshold for discriminating alcohol abusers (SDAI > 60 g/d) from social users. Similarly, Yegles et al. proposed the lower limit of 25 pg/mg for minimizing the false-positive rate when detecting an alcohol abuse through the HEtG test. On the other hand, Politi et al. investigating 21 social drinkers (SDAI < 60 g/d) and 22 heavy drinkers (SDAI >= 60 g/d) adopted a 4.0 pg/mg cutoff, which led to a sensitivity of 1.00, limiting the specificity to 0.67. To avoid any false-positive results, Kintz et al. proposed a conservative 50 pg/mg cutoff. More recently, Kharbouche et al. investigating 43 teetotalers (SDAI = 0 g), 44 low-risk drinkers (SDAI <= 30 g), and 38 at-risk drinkers (SDAI > 30 g) found that the 25 g/mg cutoff provided the best compromise between sensitivity (0.95) and specificity (0.97) for detecting at-risk drinking behaviors. Conversely, Lees et al. showed that the 30 pg/mg threshold, introduced for the first time by Bendroth et al., provided 0.58 of sensitivity and 0.86 of specificity, whereas the 45 pg/mg cutoff lowered the sensitivity to 0.52 and increased the specificity to 0.89.

Our study pooled all the individual SDAI and HEtG data available in the international published literature and performed an ROC analysis to identify the best cutoffs of the HEtG test that discriminate an alcohol ingestion of more than 60 g/d.

The pooled analysis included 8 studies, which reported paired data for 70 subjects, allowing the ROC curve calculations reported in Table 3, which shows that the 10 pg/mg cutoff provides the best sensitivity of the test (1.00) suggesting its potential utility for screening purposes, when the key point is to limit the number of false-negative results. On the other hand, the 30 pg/mg cutoff value led to the best specificity (0.97) combined with acceptable sensitivity (0.85), suggesting its applicability in a setting where the rate of false positives needs to be minimized.

The data on the diagnostic performance of the HEtG test derived from the meta-analysis (cutoff fixed at 30 pg/mg) were even better than those of the pooled analysis, with a sensitivity of 0.96 (95% CI, 0.72–1.00) and a specificity of 0.99 (95% CI, 0.92–1.00), although in the presence of a high variability explained as true heterogeneity (Fig. 5).

To the best of our knowledge, only 3 statistical evaluations by ROC analysis of HEtG concentrations in scalp hair have been described until now to determine a cutoff value for
differentiating alcohol abusers from social drinkers or abstainers. As well known, the ROC curve plots the sensitivity against the false-positive rate (ie, 1 - specificity), and each point of the curve reflects the diagnostic performance of the test for a specific cutoff value. [58]

As shown in Figure 5, the ROC curve calculated in our meta-analysis is apparently impressive, with an AUC of 0.99 (95% CI, 0.98–1.00). However, because the 95% confidence region for combined sensitivity and specificity is quite wide (ranging from 0.38 to 1.00 for sensitivity and from 0.80 to 1.00 for specificity), it must also be considered that the AUC value is an approximation, ranging from a minimum hypothetical value of 0.76 to a maximum of 0.99. With an AUC value above 0.90, the test is generally defined excellent, with a value between 0.80 and 0.90 very good, between 0.70 and 0.80 good, and with a value below 0.60 it is usually defined poor performance. In the light of the above, our estimated AUC value (0.76–0.99) identifies at least a good performing test for detecting the targeted condition (ie, SDAI > 60 g).

In 2009, the SoHT identified the 30 pg/mg HEtG value as the most useful threshold for detecting a chronic excessive alcohol consumption (ie, average intake of more than 60 g of pure ethanol per day over several months), publishing 2 consensus documents and recommendations. Our pooled analysis suggests that HEtG concentration is a promising test for confirming the targeted condition (when really present) and for excluding it when negative (SDAI < 60 g/d), even if as a consequence of the wide CI of the summary likelihood ratio point these conclusions are merely indicative (Fig. 6A).

Moreover, it has to be considered that in the clinical and/or forensic practice (ie, screening of the general population), the prevalence of the targeted condition (chronic excessive drinking) is much lower than in the meta-analyzed studies, usually being less than 10%. Following these assumptions, we have calculated that with a positive test, the posttest probability to really face an alcohol abuser will be around 80%–90%, whereas with a negative test, the probability will only be around 0.2%–0.5% (Fig. 6B). These data suggest that, at the 30 pg/mg cutoff, the HEtG test could be a useful tool for screening large populations to exclude the presence of an excessive drinking (targeted condition) but that a positive test should be confirmed by adjunctive clinical and biochemical investigations, such as carbohydrate-deficient transferrin, mean corpuscular erythrocyte volume, transaminases, gamma-glutamyltransferase, and phosphatidylethanol in blood.

Because of the fact that the apparently impressive figures for sensitivity, specificity, likelihood ratios, PPV, and NPV are derived from studies affected by a number of significant limitations (see below) and that the HEtG variability explained by the SDAI is quite limited, its use as a forensic marker still requires more in-depth validation studies.
Study Limitations

From our sample size and power calculations, given a 5%–10% theoretical prevalence of the targeted condition within the general population, a total sample size ranging from 200 subjects (including 20 heavy drinkers) in the case of a statistical power equal to 0.80 with disease prevalence of 10% to a required total sample size of 700 subjects (including 35 heavy drinkers) in the case of a statistical power equal to 0.95 with disease prevalence of 5% is required to appropriately evaluate the statistical significance for combined sensitivity and specificity values of HEtG as a marker of chronic excessive drinking, when applied to the general population. Based on the fact that, at the moment, none of the published studies exhibit such a high number of recruited subjects, we have aggregated usable data derived from single studies to reach a higher statistical consistency, still remaining, however, under the above-mentioned ideal figures. Indeed, although the number of subjects included in the pooled analysis classifies it as one of the largest single study populations available in the literature, since the introduction of the HEtG analysis, the limited number of subjects included in each of the considered studies remains a weakness and might affect the calculation of the CIs and heterogeneity.

A second weakness is the potential exposure to stressful environmental conditions and cosmetic and washing products, which could represent a source of bias because they are generally unknown to the researcher and only rarely reported in the included manuscripts. [5,13,28,61]

A third weakness is the potential inaccurate estimation of the alcohol consumption, with possible under- or overreporting of the SDAI. Specifically developed questionnaires were rarely adopted in the included studies. This may lead to a misclassification of a subject and to the calculation of unreliable cutoff values. Moreover, data collection may itself be biased by further drawbacks. The assessment of alcohol consumption may rely on either retrospective or prospective methods. Prospective alcohol self-monitoring reports clearly represent a superior alternative to retrospective self-reported methods, which are associated with a lack of compliance and more frequent missing data (ie, far memory recall in alcohol-caused brain damage and/or alcohol-induced memory lapses).17 For these reasons, we propose to refer the experimentally controlled ingestion of alcohol as DAI,18 using SDAI for all other circumstances, thus highlighting when an approximation of the reported estimation of uncontrolled alcohol assumptions has been used.

A fourth weakness is because of the presence of publication biases, consisting of small studies overestimating the effect or, alternatively, of the lack of negative/nonsignificant small studies.

A last weakness was demonstrated by a tool aimed at evaluating the methodological quality of the included studies (ie, modified Quality Assessment of Diagnostic Accuracy Studies-2), which showed that the patient selection and interpretation of the HEtG results were at high risk of bias for 22.2% of the articles, the interpretation of the reference standard (ie, self-reported ethanol
consumption, questionnaires, etc.) was at high risk of bias for 77.8% of the studies, and the patient flow/timing of analysis was at high risk of bias for 44.0% of the included studies.

**Conclusions**

The present study evidenced a good performance of HEtG as a screening test for excessive chronic drinking and a mild association between SDAI and HEtG data. The 30 pg/mg cutoff for HEtG, identified in 2009 by the SoHT, led to apparently impressive values of sensitivity, specificity, likelihood ratios, and PPV/NPV, even though a potential misclassification of an individual's drinking habit cannot be ruled out when only HEtG is used for the diagnosis.

Additionally, the present meta-analyzed data should be interpreted with caution as they are derived from studies affected by a number of relevant limitations, potentially biasing the obtained results. The sample size analyses indicate that larger and well-designed population studies are required to further validate the utilization of the HEtG marker in the forensic setting.

**REFERENCES**


FIGURE 1. Systematic search process and articles included in the analysis.

TABLE 1 - a Articles Included Into the Pooled/Meta-Analysis
### TABLE 1. (Continued) Articles Included into the Pooled/Meta-Analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Main Finding of the Study</th>
<th>Pooled Analysis</th>
<th>Meta-Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yege et al</td>
<td>Cutoff 30 pg/mg, 100%*</td>
<td>Cutoff 30 pg/mg, 100%*</td>
<td>Cutoff 30 pg/mg, 87.5%*</td>
<td>ECG is a marker of chronic excessive drinking. There seems to be weak correlation between HEG and TAI</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

Features of selected studies, investigated populations, alcohol consumption, and diagnostic efficiency of HEG analysis are reported. Data are reported as mean ± SD, or total range (5) according to the type of statistical distribution.

*Calculated as robust data.

### TABLE 1 - b Articles Included Into the Pooled/Meta-Analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Main Finding of the Study</th>
<th>Pooled Analysis</th>
<th>Meta-Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yege et al</td>
<td>Cutoff 30 pg/mg, 100%*</td>
<td>Cutoff 30 pg/mg, 100%*</td>
<td>Cutoff 30 pg/mg, 87.5%*</td>
<td>ECG is a marker of chronic excessive drinking. There seems to be weak correlation between HEG and TAI</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

Features of selected studies, investigated populations, alcohol consumption, and diagnostic efficiency of HEG analysis are reported. Data are reported as mean ± SD, or total range (5) according to the type of statistical distribution.

*Calculated as robust data.

### TABLE 1 - c Articles Not Included Into the Pooled/Meta-Analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Main Finding of the Study</th>
<th>Pooled Analysis</th>
<th>Meta-Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yege et al</td>
<td>Cutoff 30 pg/mg, 100%*</td>
<td>Cutoff 30 pg/mg, 100%*</td>
<td>Cutoff 30 pg/mg, 87.5%*</td>
<td>ECG is a marker of chronic excessive drinking. There seems to be weak correlation between HEG and TAI</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

Features of selected studies, investigated populations, alcohol consumption, and diagnostic efficiency of HEG analysis are reported. Data are reported as mean ± SD, or total range (5) according to the type of statistical distribution.

*Calculated as robust data.
TABLE 2 - b Articles Not Included Into the Pooled/Meta-Analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>Estimation of Alcohol Use</th>
<th>Methodology</th>
<th>Sample Size</th>
<th>Analytical Method</th>
<th>ETG Concentration</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hesse et al.</td>
<td>Self-reported consumption</td>
<td>Cross-sectional</td>
<td>690</td>
<td>GCMS-MS</td>
<td>10 pg/mg</td>
<td>95%</td>
</tr>
<tr>
<td>Letal et al.</td>
<td>Daily alcohol intake</td>
<td>Cross-sectional</td>
<td>662</td>
<td>GCMS-MS</td>
<td>10 pg/mg</td>
<td>95%</td>
</tr>
</tbody>
</table>

FIGURE 2. Regression parabolic line describing the mathematical relationship between HETG concentration and SDAI data (HETG = 8.1895 + 0.5234 x SDAI - 0.0003224 x SDAI^2). Parabolic continuous line: regression line. Dark (inner) dashed lines: 95% CI for the regression line. This interval includes the true regression line with a probability of 95%. Light (outer) dashed lines: 95% prediction interval for the regression curve, wider than the 95% CI. For any given value of the independent variable (estimated admitted daily alcohol intake), this interval represents the 95% probability for the values of the dependent variable (EtG concentration in hair). ETG, ethyl glucuronide.

TABLE 2 - c Articles Not Included Into the Pooled/Meta-Analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>ETG Analysis (Hair)</th>
<th>PPV</th>
<th>NPV</th>
<th>Main Finding of the Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feen et al.</td>
<td>HETG was found more accurate in terms of sensitivity and specificity with respect to MCV, AST, ALT, GGT, and CDT.</td>
<td>—</td>
<td>—</td>
<td>HETG should be interpreted with caution in renal disease.</td>
</tr>
<tr>
<td>Holm et al.</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>HETG was found more accurate in terms of sensitivity and specificity with respect to MCV, AST, ALT, GGT, and CDT.</td>
</tr>
</tbody>
</table>

Features of selected studies, investigated populations, alcohol consumption, and diagnostic efficiency of ETG analysis in hair are reported. Data are reported as mean (M) ± SE.
FIGURE 3 . Sensitivity and specificity plot as a function of the HEtG concentration. Continuous thick line represents sensitivity, with its 95% CI (continuous thin lines). Dashed thick line represents specificity, with its 95% CI (dashed thin lines).

![Graph showing sensitivity and specificity plot](image)

TABLE 3 Coordinates of the ROC Curve

<table>
<thead>
<tr>
<th>Test</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>LRB (95% CI)</th>
<th>LSR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>0.80 (0.70 - 0.90)</td>
<td>0.70 (0.60 - 0.80)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>0.70 (0.60 - 0.80)</td>
<td>0.80 (0.70 - 0.90)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>0.60 (0.50 - 0.70)</td>
<td>0.90 (0.80 - 1.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>0.50 (0.40 - 0.60)</td>
<td>1.00 (0.90 - 1.10)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: df, degrees of freedom; I2, statistical index of heterogeneity; Q, Cochrane-Q test.

FIGURE 4 . Forest plots of both sensitivity and specificity for the HEtG test with SDAI as a reference. Gray squares represent the calculated specificity or sensitivity for the test within each study; the gray line identifies the corresponding 95% CI. Abbreviations: df, degrees of freedom; I2, statistical index of heterogeneity; Q, Cochrane-Q test.

![Forest plots of sensitivity and specificity](image)
FIGURE 5. Summary ROC plot of the sensitivity and specificity for the HEtG test with SDAI as a reference. The sensitivity of the test was plotted against the specificity with inverted axis, allowing the comparison of both parameters for multiple tests. The empty circles represent each of the considered studies, and their size is proportional to the number of the included patients. The filled square is the summary estimation of the sensitivity and specificity; the dotted irregular areas around it represent the 95% confidence region (orange large dotted irregular areas) and the 95% prediction region (gray short dotted irregular areas).

FIGURE 6. Likelihood ratio scattergram. A, The summary point of likelihood ratio was obtained as a function of the mean sensitivity and specificity values. LRP and LRN (brown dotted straight lines) were settled at 28.9 and 0.16, respectively, splitting the graphical area into the following quadrants: left upper quadrant (LUQ), representing the area in which the test is useful for both confirmation and exclusion of the targeted condition; right upper quadrant (RUQ), representing the area in which the test is useful only for confirmation; left lower quadrant (LLQ), representing the area in which the test is useful only for exclusion; right lower quadrant (RLQ), representing the area in which the test is useless for both confirmation and exclusion. The brown diamond represents the summary point of LR for the HEtG test; the orthogonal solid straight lines identify the 95% CI of the estimation. Fagan nomogram for HEtG. B, The nomogram consists of a vertical axis on the left with the previous log odds, an axis in the middle displaying the log-likelihood ratio and a vertical axis on the right representing the posterior log odds. Lines were drawn from the previous probability on the left (42%) through the calculated likelihood ratios in the center (LRP = 69 and LRN = 0.04) and extended to the posterior probabilities on the right, obtaining the estimated positive and negative posterior probabilities (solid brown line and dashed blue line, respectively). For other pretest probabilities arising from different population distributions, the posttest probability can be calculated by simply drawing a line from the given pretest probability to the positive and/or negative likelihood ratio in the middle and extending it to cross the scale on the right. LRP, likelihood ratio-positive; LRN, likelihood ratio-negative.
<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Overall Statistical Power</th>
<th>Prevalence of Disease in General Population</th>
<th>Subjects With Condition (n)</th>
<th>Subjects Without Condition (n)</th>
<th>Total Subjects (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.80</td>
<td>0.98</td>
<td>≥0.95</td>
<td>&lt;0.05</td>
<td>35</td>
<td>665</td>
<td>700</td>
</tr>
<tr>
<td>D.80</td>
<td>0.98</td>
<td>≥0.95</td>
<td>&lt;0.05</td>
<td>35</td>
<td>315</td>
<td>350</td>
</tr>
<tr>
<td>D.80</td>
<td>0.98</td>
<td>≥0.99</td>
<td>&lt;0.05</td>
<td>28</td>
<td>512</td>
<td>560</td>
</tr>
<tr>
<td>D.80</td>
<td>0.98</td>
<td>≥0.99</td>
<td>&lt;0.05</td>
<td>28</td>
<td>212</td>
<td>280</td>
</tr>
<tr>
<td>D.80</td>
<td>0.98</td>
<td>≥0.80</td>
<td>&lt;0.03</td>
<td>20</td>
<td>300</td>
<td>400</td>
</tr>
<tr>
<td>D.80</td>
<td>0.98</td>
<td>≥0.80</td>
<td>&lt;0.03</td>
<td>20</td>
<td>180</td>
<td>200</td>
</tr>
</tbody>
</table>

Given the prevalence of the targeted condition in the general population, sample size (including both subjects with and without the targeted condition) required to satisfy the minimum expected conditions for sensitivity, specificity, overall statistical power, and $p$-value is reported.

TABLE 4 Sample Size and Power Calculation