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

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 (ie, 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 <=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 (ie, prevalence of the targeted condition in the investigated population). To assess the potential publication bias, the Begg–Mazumdar adjusted rank correlation analysis,48 the Egger regression of standardized effect estimates against their precision,49 the Harbord modified Egger test, and the Duval and Tweedie nonparametric analysis 50 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%51 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: $HEtG = 8.1895 + 0.5234 \times SDAI - 0.0003224 \times SDAI2$ (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 admittable 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

125.5; P < 0.00). 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,

http://links.lww.com/TDM/A83). 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.6 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.[33]

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 13 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,27 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 21 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 11 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 32 proposed a conservative 50 pg/mg cutoff. More recently, Kharbouche et al 17 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 19 showed that the 30 pg/mg threshold, introduced for the first time by Bendroth et al,35 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.13,17,25 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.58,59 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. [36–38]

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. [25,42,60]

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

1. Viel G, Boscolo-Berto R, Cecchetto G, et al.. Phosphatidylethanol in blood as a marker of chronic alcohol use. A systematic review and meta-analysis. Int J Mol Sci. 2012;13:14788–14812.

2. Wurst FM, Skipper GE, Weinmann W. Ethyl glucuronide—the direct ethanol metabolite on the threshold from science to routine use. Addiction. 2003;98(suppl 2):51–61.

3. Morini L, Politi L, Acito S, et al.. Comparison of ethyl glucuronide in hair with carbohydrate-deficient transferrin in serum as markers of chronic high levels of alcohol consumption. Forensic Sci Int. 2009;188:140–143.

4. Nalesso A, Viel G, Cecchetto G, et al.. Quantitative profiling of phosphatidylethanol molecular species in human blood by liquid chromatography high resolution mass spectrometry. J Chromatogr A. 2011;1218:8423–8431.

5. Tagliaro F, Bortolotti F, Viel G, et al.. Caveats against an improper use of hair testing to support the diagnosis of chronic excessive alcohol consumption, following the “Consensus” of the Society of Hair Testing 2009 [Forensic Science International 196 (2010) 2]. Forensic Sci Int. 2011;207:e69–e70.

6. Boscolo-Berto R, Viel G, Montisci M, et al.. Ethylglucuronide concentration in hair for detecting heavy drinking and/or abstinence. A meta-analysis. Int J Legal Med. 2013;127:611–619.

7. Kerekes I, Yegles M, Grimm U, et al.. Ethyl glucuronide determination: head hair versus non-head hair. Alcohol Alcohol. 2009;44:62–66.

8. Favretto D, Nalesso A, Frison G, et al.. A novel and an effective analytical approach for the LC-MS determination of ethyl glucuronide and ethyl sulfate in urine. Int J Legal Med. 2010;124:161–164.

9. Boscolo-Berto R, Viel G, Cecchi R, et al.. Journals publishing bio-medicolegal research in Europe. Int J Legal Med. 2012;126:129–137.

10. Ferrara SD, Bajanowski T, Cecchi R, et al.. Bio-medicolegal scientific research in Europe: a comprehensive bibliometric overview. Int J Legal Med. 2011;125:393–402.

11. Politi L, Morini L, Leone F, et al.. Ethyl glucuronide in hair: is it a reliable marker of chronic high levels of alcohol consumption? Addiction. 2006;101:1408–1412.

12. Pragst F, Rothe M, Moench B, et al.. Combined use of fatty acid ethyl esters and ethyl glucuronide in hair for diagnosis of alcohol abuse: interpretation and advantages. Forensic Sci Int. 2010;196:101–110.

13. Morini L, Politi L, Polettini A. Ethyl glucuronide in hair. A sensitive and specific marker of chronic heavy drinking. Addiction. 2009;104:915–920.

14. Viel G, Boscolo-Berto R, Cecchi R, et al.. Bio-medicolegal scientific research in Europe. A country-based analysis. Int J Legal Med. 2011;125:717–725.

15. Pirro V, Di Corcia D, Pellegrino S, et al.. A study of distribution of ethyl glucuronide in different keratin matrices. Forensic Sci Int. 2011;210:271–277.

16. Janda I, Weinmann W, Kuehnle T, et al.. Determination of ethyl glucuronide in human hair by SPE and LC-MS/MS. Forensic Sci Int. 2002;128:59–65.

17. Kharbouche H, Faouzi M, Sanchez N, et al.. Diagnostic performance of ethyl glucuronide in hair for the investigation of alcohol drinking behavior: a comparison with traditional biomarkers. Int J Legal Med. 2012;126:243–250.

18. Kronstrand R, Brinkhagen L, Nystrom FH. Ethyl glucuronide in human hair after daily consumption of 16 or 32 g of ethanol for 3 months. Forensic Sci Int. 2012;215:51–55.

19. Lees R, Kingston R, Williams TM, et al.. Comparison of ethyl glucuronide in hair with self-reported alcohol consumption. Alcohol Alcohol. 2012;47:267–272.

20. Shi Y, Shen B, Xiang P, et al.. Determination of ethyl glucuronide in hair samples of Chinese people by protein precipitation (PPT) and large volume injection-gas chromatography-tandem mass spectrometry (LVI-GC/MS/MS). J Chromatogr B Analyt Technol Biomed Life Sci. 2010;878:3161–3166.

21. Yegles M, Labarthe A, Auwärter V, et al.. Comparison of ethyl glucuronide and fatty acid ethyl ester concentrations in hair of alcoholics, social drinkers and teetotallers. Forensic Sci Int. 2004;145:167–173.

22. Lamoureux F, Gaulier JM, Sauvage FL, et al.. Determination of ethyl-glucuronide in hair for heavy drinking detection using liquid chromatography-tandem mass spectrometry following solid-phase extraction. Anal Bioanal Chem. 2009;394:1895–1901.

23. Schrader J, Rothe M, Pragst F. Ethyl glucuronide concentrations in beard hair after a single alcohol dose: evidence for incorporation in hair root. Int J Legal Med. 2012;126:791–799.

24. Agius R, Nadulski T, Kahl HG, et al.. Ethyl glucuronide in hair—a highly effective test for the monitoring of alcohol consumption. Forensic Sci Int. 2012;218:10–14.

25. Pirro V, Valente V, Oliveri P, et al.. Chemometric evaluation of nine alcohol biomarkers in a large population of clinically-classified subjects: pre-eminence of ethyl glucuronide concentration in hair for confirmatory classification. Anal Bioanal Chem. 2011;401:2153–2164.

26. Alt A, Janda I, Seidl S, et al.. Determination of ethyl glucuronide in hair samples. Alcohol Alcohol. 2000;35:313–314.

27. Appenzeller BM, Agirman R, Neuberg P, et al.. Segmental determination of ethyl glucuronide in hair: a pilot study. Forensic Sci Int. 2007;173:87–92.

28. Morini L, Zucchella A, Polettini A, et al.. Effect of bleaching on ethyl glucuronide in hair: an in vitro experiment. Forensic Sci Int. 2010;198:23–27.

29. Suesse S, Pragst F, Mieczkowski T, et al.. Practical experiences in application of hair fatty acid ethyl esters and ethyl glucuronide for detection of chronic alcohol abuse in forensic cases. Forensic Sci Int. 2012;218:82–91.

30. Gareri J, Appenzeller B, Walasek P, et al.. Impact of hair-care products on FAEE hair concentrations in substance abuse monitoring. Anal Bioanal Chem. 2011;400:183–188.

31. Kerekes I, Yegles M. Coloring, bleaching, and perming: influence on EtG content in hair. Ther Drug Monit. 2013;35:527–529.

32. Kintz P, Villain M, Vallet E, et al.. Ethyl glucuronide: unusual distribution between head hair and pubic hair. Forensic Sci Int. 2008;176:87–90.

33. Pragst F, Yegles M. Determination of fatty acid ethyl esters (FAEE) and ethyl glucuronide (EtG) in hair: a promising way for retrospective detection of alcohol abuse during pregnancy? Ther Drug Monit. 2008;30:255–263.

34. Albermann ME, Musshoff F, Madea B. A fully validated high-performance liquid chromatography-tandem mass spectrometry method for the determination of ethyl glucuronide in hair for the proof of strict alcohol abstinence. Anal Bioanal Chem. 2010;396:2441–2447.

35. Bendroth P, Kronstrand R, Helander A, et al.. Comparison of ethyl glucuronide in hair with phosphatidylethanol in whole blood as post-mortem markers of alcohol abuse. Forensic Sci Int. 2008;176:76–81.

36. Kintz P. Consensus of the Society of Hair Testing on hair testing for chronic excessive alcohol consumption 2009. Forensic Sci Int. 2010;196:2.

37. Kintz P. Consensus of the Society of Hair Testing on hair testing for chronic excessive alcohol consumption 2011. Forensic Sci Int. 2012;218:2.

38. Aronoff GR, Berns JS, Brier ME, et al.. Drug Prescribing in Renal Failure: Dosing Guidelines for Adults. 4th ed. Philadelphia, PA: American College of Physicians; 1999.]

39. Crunelle CL, Yegles M, Nuijs AL, et al.. Hair ethyl glucuronide levels as a marker for alcohol use and abuse: a review of the current state of the art. Drug Alcohol Depend. 2014;134:1–11.

40. Pragst F. Interpretation problems in a forensic case of abstinence determination using alcohol markers in hair. Forensic Sci Int. 2012;217:e4–e7.

41. Pirro V, Di Corcia D, Seganti F, et al.. Determination of ethyl glucuronide levels in hair for the assessment of alcohol abstinence. Forensic Sci Int. 2013;232:229–236.

42. Pirro V, Oliveri P, Sciutteri B, et al.. Multivariate strategies for screening evaluation of harmful drinking. Bioanalysis. 2013;5:687–699.

43. Boscolo-Berto R, Viel G, Montagnese S, et al.. Narcolepsy and effectiveness of gamma-hydroxybutyrate (GHB): a systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev. 2012;16:431–443.

44. Boscolo-Berto R. Clinical testing and evidence-based medicine: when the absence of evidence doesn't mean evidence of absence. G Ital Nefrol. 2009;26:417.

45. Higgins JP, Altman DG, Gotzsche PC, et al.. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.

46. Reitsma JB, Glas AS, Rutjes AW, et al.. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58:982–990.

47. Arends LR, Hamza TH, van Houwelingen JC, et al.. Bivariate random effects meta-analysis of ROC curves. Med Decis Making. 2008;28:621–638.

48. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics. 1994;50:1088–1101.

49. Egger M, Davey Smith G, Schneider M, et al.. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315:629–634.

50. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000;56:455–463.

51. Ferrara SD, Favretto D, Montisci M, et al.. Prevalence of Alcohol and Other Psychoactive Substances in Drivers in General Traffic. Country Report Italy. DRUID Deliverable D 2.2.3. 2011:108–123.

52. Li J, Fine J. On sample size for sensitivity and specificity in prospective diagnostic accuracy studies. Stat Med. 2004;23:2537–2550.

53. Obuchowski NA, Zhou XH. Prospective studies of diagnostic test accuracy when disease prevalence is low. Biostatistics. 2002;3:477–492.

54. Zhou XH, Obuchowski NA, McClish DK. Statistical Methods in Diagnostic Medicine. New York, NY: Wiley-Blackwell; 2002.

55. Moher D, Liberati A, Tetzlaff J, et al.. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6:e1000097.

56. Whiting PF, Rutjes AW, Westwood ME, et al.. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529–536.

57. Høiseth G, Morini L, Ganss R, Nordal K, Morland J. Higher levels of hair ethyl glucuronide in patients with decreased kidney function. Alcohol Clin Exp Res. 2013;37Suppl1:E14–E16.

58. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29–36.

59. Choi BC. Slopes of a receiver operating characteristic curve and likelihood ratios for a diagnostic test. Am J Epidemiol. 1998;148:1127–1132.

60. Vincenti M, Salomone A, Pirro V. How has screening of harmful drinking changed over the years? Bioanalysis. 2013;5:2981–2983.

61. Pragst F, Balikova MA. State of the art in hair analysis for detection of drug and alcohol abuse. Clin Chim Acta. 2006;370:17–49.

Image gallery

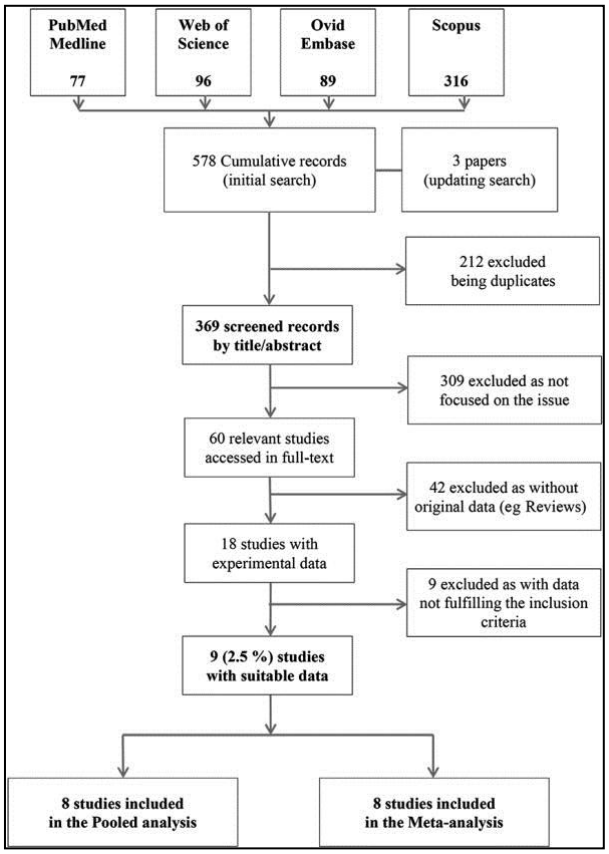

FIGURE 1 . Systematic search process and articles included in the analysis.

	Features of the Study			Investigated Population		
Study	Main Aim of Interest	Inclusion Criteria	Exclusion Criteria	No. Subjects	Experimental Setting	Subject Stratification
Appenzeller et al[27]	Compare EtG hair concentrations with the self-reported drinking history of alcoholics under detoxification	Age ≥ 18 yrs informed consent	—	15 subjects (11 M/4 F)	Inpatients	—
Kerekes et al[7]	Compare the concentrations of EtG	Autopsy cases and healthy volunteers	—	32 subjects (25 with self-reported alcohol consumption)	Forensic autopsy cases and healthy volunteers	8 teetotalers, EDAI < 10 g/d 13 social drinkers, EDAI < 40 g/d 4 heavy drinkers, EDAI > 60 g/d
Lamoureux et al[22]	Validate an LC-ESI-MS/MS method for HEtG determination in 30 mg of hair samples	Heavy drinkers and social drinkers	Head hair < 3 cm	19 subjects (16 M/3 F)	Inpatients recruited at the emergency department, healthy volunteers	7 social drinkers (0–20 g/d) 12 heavy drinkers (>50 g/d for several wk)
Morini et al[3]	Compare the diagnostic efficiency of EtG in hair and CDT for diagnosing a chronic alcohol abuse	Volunteers	Head hair < 3 cm	86 subjects (48 M/38 F)	Inpatients and outpatients	24 subjects with EDAI 0–59 g/d 62 subjects with EDAI >60 g/d
Morini et al[28]	Evaluate the effect of hair bleaching on [EtG]	Volunteers	<3 cm hair length	30 subjects (19 M/11 F)	Inpatients and outpatients	5 teetotalers 4 social drinkers (30–60 g EDAI) 21 heavy drinkers (EDAI ≥ 60 g/d)
Pirro et al[15]	Evaluate EtG levels in head, pubic, and axillary hair	Healthy volunteers from an alcohol abuse treatment service	Uncertain estimation of the daily alcohol intake hair length < 3 cm	32 healthy volunteers (25 M/7 F)	Inpatients	9 nondrinkers 11 social drinkers 12 heavy drinkers
Politi et al[13]	Correlate the concentration of EtG in hair with the EDAI	Inpatients of an alcohol abuse treatment service and healthy volunteers	Hair length < 3 cm	50 subjects	Inpatients and healthy volunteers	21 social drinkers (EDAI < 60 g/d) 22 heavy drinkers (EDAI > 60 g/d) 7 teetotalers
Shi et al[29]	Develop and validate a sensitive LVI-GC-MS/MS method for HEtG	Healthy volunteers	—	21 subjects (15 M/6 F)	Healthy volunteers	6 teetotalers (children), EDAI = 0 g/d 4 social drinkers (EDAI ≤ 30 g/d) 11 heavy drinkers (EDAI > 60 g/d)
Yegles et al[21]	To compare EtG and FAEE concentration in hair with EDAI	Inpatients of an alcohol abuse treatment service, healthy volunteers	—	17 subjects	Inpatients and healthy volunteers	3 teetotalers, EDAI = 0 g/d 4 social drinkers (EDAI ≤ 20 g/d) 10 heavy drinkers (EDAI > 60 g/d)

	Alcohol Consumption		HEtG Analysis			
Study	Estimation of Alcohol Use	SDAI Before Test, g	Hair Length	Analytical Method LOD, LOQ, pg/mg	HEtG Concentration, pg/mg	Sensitivity
Appenzeller et al[27]	Self-reported MALT (>1 mo)	296 ± 143	0.5–8 cm proximal	GC/MS, LOD 2, LOQ 4	113.3 ± 80.9	Cutoff 30 pg/mg, 86.7%
Kerekes et al[7]	Self-reported consumption (>1 mo)	0 25.4 ± 8.0 72.5 ± 12.6	6 cm proximal	GC/MS, LOD 0.7, LOQ 2.3	2.1 ± 2.6 13.5 ± 11.2 107.3 ± 81.7	Cutoff 30 pg/mg, EDAI > 60/d, 100%*

TABLE 1 -a Articles Included Into the Pooled/Meta-Analysis

TABLE 1. (*Continued*) Articles Included Into the Pooled/Meta-Analysis

	Alcohol Consumption		HEtG Analysis			
Study	Estimation of Alcohol Use	SDAI Before Test, g	Hair Length	Analytical Method LOD, LOQ, pg/mg	HEtG Concentration, pg/mg	Sensitivity
Lamoureux et al[22]	Self-reported consumption (>1 mo)	6.6 ± 3.2 >50	—	LC/MS-MS, LOD 4, LOQ 10	<LOQ 220 ± 203.9	Cutoff 30 pg/mg, 100%*
Morini et al[5]	TLFB (3 mo)	Male (n = 48) 181.0 ± 149.6 Female (n = 38) 135.8 ± 125.4	3 cm proximal	LC/MS-MS, LOD 2, LOQ 3		Cutoff 27 pg/mg, 98.4%
Morini et al[28]	Self-reported consumption (3 mo)	0 30–60$ ≥60	3 cm proximal	LC/MS-MS, LOD 2, LOQ 3	<LOQ 12.9 ± 3.1 71.9 ± 32.3	Cutoff 27 pg/mg, 100%*
Pirro et al[15]	Self-reported consumption (6 mo)	0 <60 >60	1–6 cm proximal	LC/MS-MS, LOD 3, LOQ 10	<LOQ 13.5 ± 12.1 202.1 ± 166.6	Cutoff 30 pg/mg, 100%*
Politi et al[11]	Self-reported consumption (6 mo)	25.5 ± 18.9 >60 0	3–5 cm proximal	LC/MS-MS, LOD 2, LOQ 3	<LOQ 5.1 ± 9 462.8 ± 93.8	Cutoff 30 pg/mg, 52%*
Shi et al[29]	Self-reported consumption (3 mo)	0 19.8 ± 11.3 126.2 ± 45.5	5 cm proximal	LVI-GC/MS/MS, LOD 5, LOQ 10	<LOQ 2.8 ± 5.5 30.4 ± 22.3	Cutoff 30 pg/mg, 36.4%*
Yegles et al[21]	Self-reported consumption (3 mo)	0 ≤20 218.0 ± 87.9	3–10 cm proximal	GC/MS, MS, LOD 2, LOQ 4	<LOQ <LOQ 159.5 ± 138.5	Cutoff 30 pg/mg, 95.2%*

	HEtG Analysis				Performed Analysis	
Study	Specificity	PPV	NPV	Main Finding of the Study	Pooled Analysis	Meta-Analysis
Appenzeller et al[27]	—	Cutoff 30 pg/mg, 100%*	—	HEtG correlates with alcohol intake	✓	
Kerekes et al[7]	Cutoff 30 pg/mg, EDAI > 60/d, 95.2%*	Cutoff 30 pg/mg, EDAI > 60/d, 80%*	Cutoff 30 pg/mg, EDAI > 60/d, 100%*	Interpretation of nonhead HEtG concentrations remains controversial	✓	✓
Lamoureux et al[22]	Cutoff 30 pg/mg, 100%*	Cutoff 30 pg/mg, 100%*	Cutoff 30 pg/mg, 100%*	EtG ≥ 50 pg/mg is linked to heavy alcohol consumption. A 10 pg/mg cutoff value might also perhaps be proposed	✓	✓
Morini et al[5]	Cutoff 27 pg/mg, 95.4%	Cutoff 27 pg/mg, 98.4%*	Cutoff 27 pg/mg, 95.4%*	HEtG concentrations can differentiate heavy alcohol consumption with increased sensitivity with respect to CDT		✓
Morini et al[28]	Cutoff 27 pg/mg, 100%*	Cutoff 27 pg/mg, 100%*	Cutoff 27 pg/mg, 100%*	Hair bleaching may cause false negatives	✓	✓
Pirro et al[15]	Cutoff 30 pg/mg, 100%*	Cutoff 30 pg/mg, 100%*	Cutoff 30 pg/mg, 100%*	EtG is an objective marker of chronic alcohol abuse	✓	✓
Politi et al[11]	Cutoff 30 pg/mg, 100%*	Cutoff 30 pg/mg, 100%*	Cutoff 30 pg/mg, 60%*	EtG is an objective marker of alcohol abuse. Best cutoff identified at 5 pg/mg	✓	✓
Shi et al[29]	Cutoff 30 pg/mg, 100%*	Cutoff 30 pg/mg, 100%*	Cutoff 30 pg/mg, 6 mo EDAI > 60/d, 58.8%*	EtG is an objective marker of alcohol abuse. Best cutoff identified at 20 pg/mg. Caution when hair color or cosmetics are used	✓	✓

(*continued on next page*)

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

TABLE 1. (*Continued*) Articles Included Into the Pooled/Meta-Analysis

	HEtG Analysis				Performed Analysis	
Study	Specificity	PPV	NPV	Main Finding of the Study	Pooled Analysis	Meta-Analysis
Yegles et al[21]	Cutoff 30 pg/mg, 100%*	Cutoff 30 pg/mg, 100%*	Cutoff 30 pg/mg, 87.5%*	EtG is a marker of chronic excessive drinking. There seems to be weak correlation between HEtG and EDAI	✓	✓

Features of selected studies, investigated populations, alcohol consumption, and diagnostic efficiency of HEtG analysis are reported. Data are reported as mean ± SD, or total range ($) according to the type of statistical distribution.
*Calculated on published data.

TABLE 1 -c Articles Included Into the Pooled/Meta-Analysis

	Features of the Study			Investigated Population		
Study	Main Aim of Interest	Inclusion Criteria	Exclusion Criteria	No. Subjects	Experimental Setting	Subjects Stratification
Alt et al[26]	Test EtG concentration in hair	Alcohol dependence diagnosed after death (information from police/family practitioner) or during life (*ICD-10* criteria)	—	15 subjects	Inpatients under detoxification and healthy volunteers	5 teetotalers (children), 6 social drinkers, EDAI < 20 g/d 4 living alcohol dependents, EDAI > 60 g/d
Janda et al[16]	Develop and validate a solid phase extraction procedure and LC-MS method for the analysis of EtG in hair	Alcohol dependence diagnosed after death (information from police/family) or during life	—	70 subjects	Inpatients under detoxification and healthy volunteers	5 teetotalers (children), 5 social drinkers, EDAI < 30 g/d 60 living alcohol dependents, EDAI > 60 g/d
Kharbouche et al[13]	Assessment of HEtG as a marker of alcohol consumption	Age > 18 yrs	Cognitive impairment, pregnancy, intravenous drug abuse	125 subjects (69 F, 56 M)	Healthy volunteers and outpatients	43 teetotalers, EDAI = 0 44 low-risk drinkers, ≤20 g/d F, ≤30 g/d M 38 at-risk drinkers, >20 g/d F, >30 g/d M
Kintz et al[32]	Compare pubic to head hair for EtG analysis	Healthy volunteers, Age 22–46 yrs, EDAI < 50 g/d	Head or pubic hair < 3 cm	8 subjects (4 M/4 F)	Healthy volunteers	Social drinkers
Kronstrand et al[18]	Evaluate HEtG levels in social drinkers and ability to differentiate them from abstainers	Randomized open controlled study	—	44 volunteers (12 M, 32 F)	Healthy volunteers	Abstainers (n = 22) Females (16 g/d) (n = 14) Males (32 g/d) (n = 7)
Lees et al[19]	Evaluate sensitivity and specificity of EtG in hair as a marker of chronic alcohol use	Inpatient alcoholics (*ICD-10* F10.25 criteria) or healthy volunteers	Head hair < 3 cm	100 subjects	Inpatients, healthy volunteers	Lower risk (8–192 g/wk; 1–24 g/d) Increasing risk (176–400 g/wk; 25–57 g/d) High risk (>400 g/wk; >57 g/d)
Morini et al[33]	Diagnostic efficiency of EtG in hair as a marker of chronic alcohol use and its correlation with EDAI	Volunteers reporting their alcohol intake over the last 2 wk and 3 mo	<3 cm hair length	98 subjects (56 M/42 F)	Healthy volunteers	EDAI 0–59 g/d (n = 23) (12 M/11 F) EDAI > 60 g/d (n = 75) (44 M/31 F)
Pirro et al[15]	Evaluate HEtG as a confirmatory test for the diagnosis of chronic alcohol abuse in forensic and clinical practice	Inpatients of an alcohol abuse treatment service (*DSM-IV* and *ICD-10* criteria)	Uncertain estimation of the daily alcohol intake, hair length < 3 cm	175 subjects classified into heavy, social, or nondrinking categories (145 M/30 F)	Inpatients	65 nondrinkers (54 M/11 F) 51 social drinkers (46 M/5 F) 59 active heavy drinkers (49 M/10 F)
Høiseth et al[17]	Investigate HEtG in patients with decreased kidney function	Nondialyzed renal disease patients (3 criteria) with moderate alcohol consumption	—	12	Inpatients	12 social drinkers with renal disease (8 M/4 F)

	Alcohol Consumption		EtG Analysis (Hair)			
Study	Estimation of Alcohol Use	SDAI Before Test (g)	Hair Length	Analytical Method LOD–LOQ Cutoff*	HEtG Concentration, pg/mg	Sensitivity
Alt et al[26]	Self-reported consumption for alcoholics (>60 g/d), for social drinkers <20 g/d and for teetotalers 0 g/d)	—	2–8 cm proximal	GC/MS, LOD not reported	<LOQ <LOQ 260.3 ± 124.4	Cutoff 30 pg/mg, 6 mo EDAI > 60/d, 90.5%*

(*continued on next page*)

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

TABLE 2. (*Continued*) Articles Not Included Into the Pooled/Meta-Analysis

Study	Alcohol Consumption: Estimation of Alcohol Use	Alcohol Consumption: SDAI Before Test (g)	EtG Analysis (Hair): Hair Length	EtG Analysis (Hair): Analytical Method LOD–LOQ Cutoff*	EtG Concentration, pg/mg	Sensitivity
Janda et al[16]	Self-reported consumption for alcoholics (>60 g/d), for social drinkers (<30 g/d) and for teetotalers (0 g/d)	—	1–33 cm proximal	LC-MS/MS, LOD, 51 pg/mg, LOQ, 102 pg/mg	<LOD <LOD (4)–55 136.2 ± 225.9	Cutoff 30 pg/mg, EDAI > 60/d, 60.9%*
Kharbouche et al[17]	Daily alcohol self-monitoring log over 3 mo	0 (IQR 0) 8.7 ± 7.2 (IQR 3.1–11.3) 64.4 ± 34.5 (IQR 36.1–88.0)	3 cm proximal	GC/MS-MS, LOD, 2 pg/mg, LOQ, 4 pg/mg	0.6 ± 2.1 4.9 ± 6.2 88.3 ± 203.9	Cutoff 25 pg/mg, 95% Cutoff 30 pg/mg, 81%
Kintz et al[32]	Self-reported consumption	<50	6 cm proximal	LC/MS-MS, LOD, 2 pg/mg, LOQ, 10 pg/mg	<LOQ	—
Kronstrand et al[18]	Drinking experiment, controlled alcohol assumption	Abstainers, 0 g/d Social drinkers, 16 g/d F Social drinkers, 32 g/d M	2 cm proximal	LC/MS-MS, LOQ, 2 pg/mg	<LOQ 2–3 * 5–11†	—
Lees et al[19]	Self-reported consumption, (3 wk) and C-AUDIT	1–24$ 25–57$ >57	1–3 cm proximal	GC/MS-MS, LOD, 30 pg/mg, LOQ, 45 pg/mg	— — —	Cutoff 30, 26%; 45, 23% Cutoff 30, 53%; 45, 50% Cutoff 30, 58%; 45, 52%
Morini et al[15]	TLFB (14 and 90 d)	0–59$ 60–119$ (n = 25) 120–179$ (n = 18) >180 (n = 32)	3 cm proximal	LC/MS-MS, LOD, 2 pg/mg, LOQ, 3 pg/mg	9.6 ± 20.7 124.8 ± 176* 198.7 ± 181.4† 254.5 ± 345.3‡	Cutoff 27 pg/mg, 3 mo EDAI, 92% Cutoff 27 pg/mg, EDAI 2 wk, 92%
Pirro et al[15]	AUDIT self-reported consumption (3 mo)	0 <60 >60	6 cm proximal	LC/MS-MS, LOD, 3 pg/mg	3.5 ± 1.2 8.0 ± 9.2 191.0 ± 173.0	Cutoff 30 pg/mg, 3 mo EDAI > 60/d, 98.2%*
Høiseth et al[57]	TLFB (3 mo)	0.4–12$	3 cm proximal	LC/MS-MS, LOD, 2 pg/mg, LOQ, 3 pg/mg	56.6 ± 45.1	—

Study	EtG Analysis (Hair): Specificity	PPV	NPV	Main Finding of the Study
Alt et al[26]	Cutoff 30 pg/mg, 6 mo EDAI > 60/d, 100%*	Cutoff 30 pg/mg, 6 mo EDAI > 60/d, 100%*	Cutoff 30 pg/mg, 6 mo EDAI > 60/d, 84.6%*	With positive EtG in hair, alcohol consumption has to be strongly assumed. With negative EtG, alcohol consumption cannot be excluded
Janda et al[16]	Cutoff 30 pg/mg, EDAI > 60/d, 90%*	Cutoff 30 pg/mg, EDAI > 60/d, 98.1%*	Cutoff 30 pg/mg, EDAI > 60/d, 20.9%*	Heavy alcohol consumption may be but not necessarily has to be detectable by EtG hair analysis
Kharbouche et al[17]	Cutoff 25 pg/mg, 97% Cutoff 30 pg/mg, 97%	Cutoff 25 pg/mg, 88% —	Cutoff 25 pg/mg, 99% —	Reliable HEtG cutoff values for identifying at-risk drinking behaviors have been detected
Kintz et al[32]	Cutoff 10 pg/mg, 100%	—	—	Head and pubic hair have different incorporation rates
Kronstrand et al[18]	Cutoff 30 pg/mg, 100%*	—	Cutoff 30 pg/mg, 100%*	No optimal differentiation between abstinence and social drinking
Lees et al[19]	Cutoff 30, 95%; 45, 100% Cutoff 30, 91%; 45, 94% Cutoff 30, 86%; 45, 89%	Cutoff 30, 95%; 45, 100% Cutoff 30, 73%; 45, 79% Cutoff 30, 50%; 45 52%	Cutoff 30, 23%; 45, 23% Cutoff 30, 82%; 45, 81% Cutoff 30, 90%; 45, 89%	HEtG conferred no advantages over reported sensitivity and specificity of CDT
Morini et al[15]	Cutoff 27 pg/mg, 3 mo EDAI, 96% Cutoff 27 pg/mg, EDAI 2 wk, 92%	—	—	A cutoff of 27 pg/mg provided the best differentiation between heavy and social drinkers

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

TABLE 2. (*Continued*) Articles Not Included Into the Pooled/Meta-Analysis

Study	EtG Analysis (Hair): Specificity	PPV	NPV	Main Finding of the Study
Pirro et al[15]	Cutoff 30 pg/mg, 3 mo EDAI > 60/d, 97.4%*	—	—	HEtG was found more accurate in terms of sensitivity and specificity with respect to MCV, AST, ALT, GGT, and CDT
Høiseth et al[57]	Cutoff 30 pg/mg, 3 mo EDAI > 60/d, 75.0%*	—	Cutoff 30 pg/mg, 100%*	HEtG should be interpreted with caution in renal disease

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

*Calculated on published data.

ALT, Alanine Aminotransferase; AST, Aspartate Aminotransferase; AUDIT, Alcohol Use Disorders Identification Test; CDT, carbohydrate-deficient transferrin; *DSM-IV, Diagnostic and Statistical Manual of mental disorders IV*; EDAI, Estimated Daily Alcohol Intake; F, Female; GC, gas chromatography; GGT, Gamma-Glutamyl Transpeptidase; *ICD, International Classification of Diseases*; IQR, interquartile range; LC, liquid chromatography; LOD, limit of detection; M, Male; MCV, mean corpuscular erythrocyte volume;

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

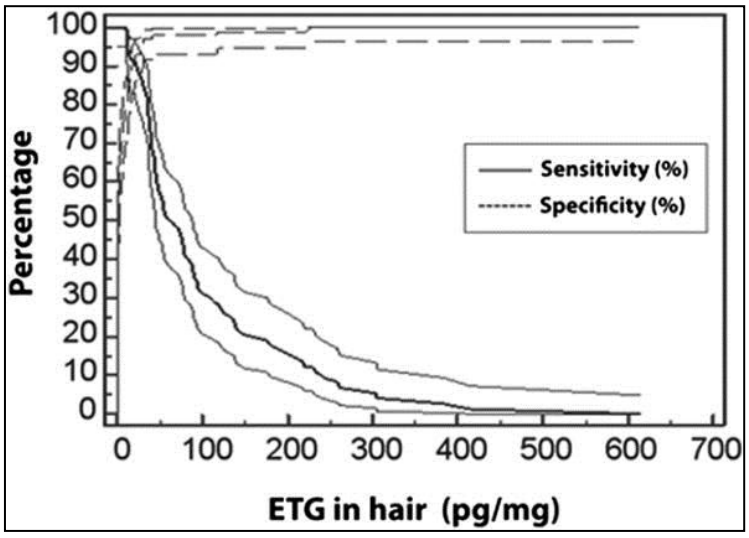

FIGURE 2 . Regression parabolic line describing the mathematical relationship between HEtG concentration and SDAI data (HEtG = 8.1895 + 0.5234 x SDAI - 0.0003224 x SDAI2). Parabolic continuous line: regression line. Dark (inner) dashed lines: 95% CI for the regression line. This interval includes the 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.

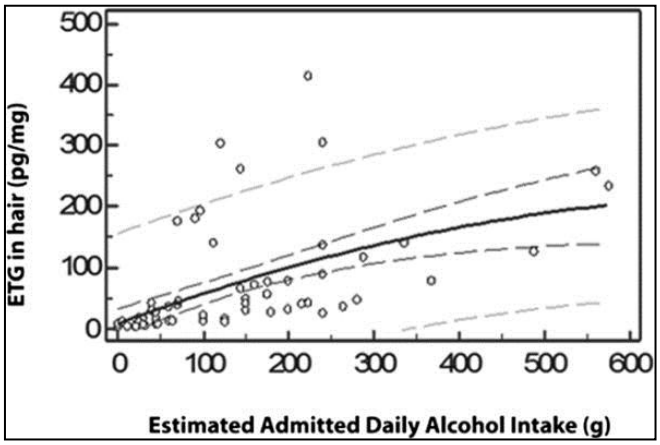

FIGURE 3 . Sensitivity and specificity plot as a function of the HEtG concentration. Continuous thick line represents sensitivity, with its 95% CI (continuous thin lines). Dashed thick line represents specificity, with its 95% CI (dashed thin lines).

EtG Cutoff, pg/mg	Sensitivity (95% CI), %	Specificity (95% CI), %	LRP (95% CI)	LRN (95% CI)
>10	100.0 (95.1–100.0)	79.4 (70.3–86.8)	4.9 (4.4–5.4)	0.00 (0.00–0.00)
>11	98.6 (92.6–100.0)	80.4 (71.4–87.6)	5.0 (4.6–5.6)	0.02 (0.01–0.10)
>12	94.5 (86.6–98.5)	84.3 (75.8–90.8)	6.0 (5.5–6.7)	0.07 (0.02–0.20)
>14	93.2 (84.7–97.7)	86.3 (78.0–92.3)	6.8 (6.1–7.5)	0.08 (0.03–0.20)
>15	91.8 (83.0–96.9)	87.3 (79.2–93.0)	7.2 (6.5–8.0)	0.09 (0.04–0.20)
>16	91.8 (83.0–96.9)	88.2 (80.4–93.8)	7.8 (7.1–8.6)	0.09 (0.04–0.20)
>17	91.8 (83.0–96.9)	90.2 (82.7–95.2)	9.4 (8.5–10.3)	0.09 (0.03–0.20)
>20	91.8 (83.0–96.9)	91.2 (83.9–95.9)	10.4 (9.5–11.4)	0.09 (0.03–0.20)
>22	90.4 (81.2–96.1)	93.1 (86.4–97.2)	13.2 (12.0–14.4)	0.10 (0.04–0.30)
>24	89.0 (79.5–95.1)	93.1 (86.4–97.2)	13.0 (11.8–14.3)	0.12 (0.04–0.30)
>26	87.7 (77.9–94.2)	93.1 (86.4–97.2)	12.8 (11.5–14.1)	0.13 (0.05–0.30)
>29	86.3 (76.2–93.2)	95.1 (88.9–98.4)	17.6 (15.9–19.5)	0.14 (0.05–0.40)
>30	84.9 (74.6–92.2)	97.1 (91.6–99.4)	28.9 (26.1–32.0)	0.16 (0.04–0.50)
>35	82.2 (71.5–90.2)	97.1 (91.6–99.4)	27.9 (25.0–31.3)	0.18 (0.05–0.60)
>36	78.1 (66.9–86.9)	97.1 (91.6–99.4)	26.6 (23.4–30.1)	0.23 (0.07–0.70)
>37	76.7 (65.4–85.8)	97.1 (91.6–99.4)	26.1 (22.9–29.7)	0.24 (0.07–0.80)
>38	74.0 (62.4–83.5)	97.1 (91.6–99.4)	25.2 (21.9–28.9)	0.27 (0.08–0.90)
>39	72.6 (60.9–82.4)	97.1 (91.6–99.4)	24.7 (21.4–28.5)	0.28 (0.09–0.90)
>40	68.5 (56.6–78.9)	97.1 (91.6–99.4)	23.3 (19.9–27.3)	0.32 (0.10–1.00)
>41	68.5 (56.6–78.9)	98.0 (93.1–99.8)	34.9 (29.8–40.9)	0.32 (0.08–1.30)
>42	67.1 (55.1–77.7)	98.0 (93.1–99.8)	34.2 (29.1–40.3)	0.34 (0.08–1.40)
>43	65.8 (53.7–76.5)	98.0 (93.1–99.8)	33.5 (28.4–39.7)	0.35 (0.09–1.40)
>44	63.1 (50.9–74.0)	98.0 (93.1–99.8)	32.1 (26.9–38.4)	0.38 (0.09–1.50)
>46	60.3 (48.1–71.5)	98.0 (93.1–99.8)	30.7 (25.5–37.1)	0.41 (0.10–1.60)
>48	58.9 (46.8–70.3)	98.0 (93.1–99.8)	30.0 (24.8–36.5)	0.42 (0.10–1.70)
>49	57.5 (45.4–69.0)	98.0 (93.1–99.8)	29.3 (24.0–35.8)	0.43 (0.10–1.80)

Coordinates of the ROC curve. The sensitivity is the probability that a test will give a positive result when the disease is present (true-positive rate). The specificity is the probability that a test will give a negative result when the disease is absent (true-negative rate). The positive likelihood ratio is the ratio between the probability of a positive test result given the presence of the disease and the probability of a positive test result given the absence of the disease. The negative likelihood ratio is the ratio between the probability of a negative test result given the presence of the disease and the probability of a negative test result given the absence of the disease. In the present study, the term "disease" has to be interpreted as "chronic excessive drinking."

LRP, likelihood ratio—positive; LRN, likelihood ratio—negative.

TABLE 3 Coordinates of the ROC Curve

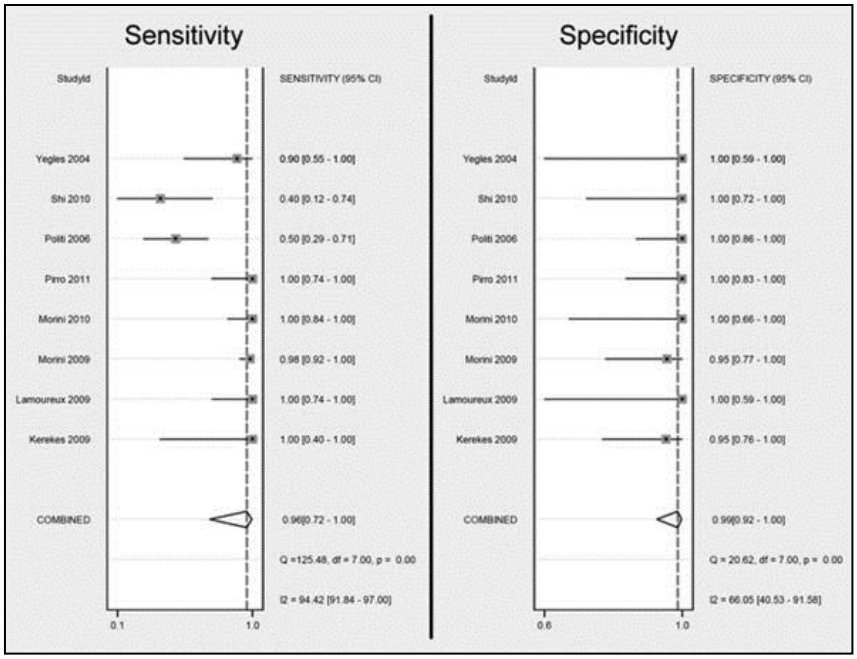

FIGURE 4 . Forest plots of both sensitivity and specificity for the HEtG test with SDAI as a reference. Gray squares represent the calculated specificity or sensitivity for the test within each study; the gray line identifies the corresponding 95% CI. Abbreviations: df, degrees of freedom; I2, statistical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

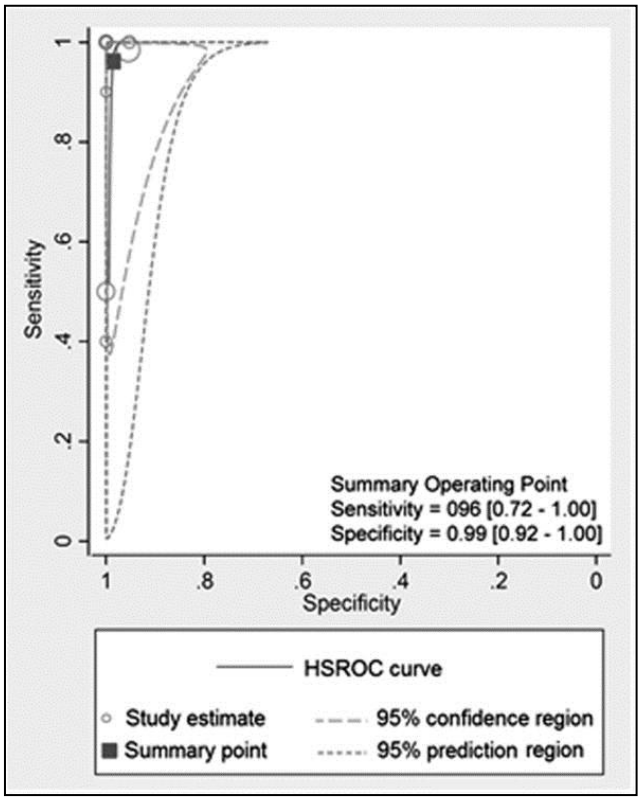

FIGURE 5 . Summary ROC plot of the sensitivity and specificity for the HEtG test with SDAI as a reference. The sensitivity of the test was plotted against the specificity with inverted axis, allowing the comparison of both parameters for multiple tests. The empty circles represent each of the considered 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).

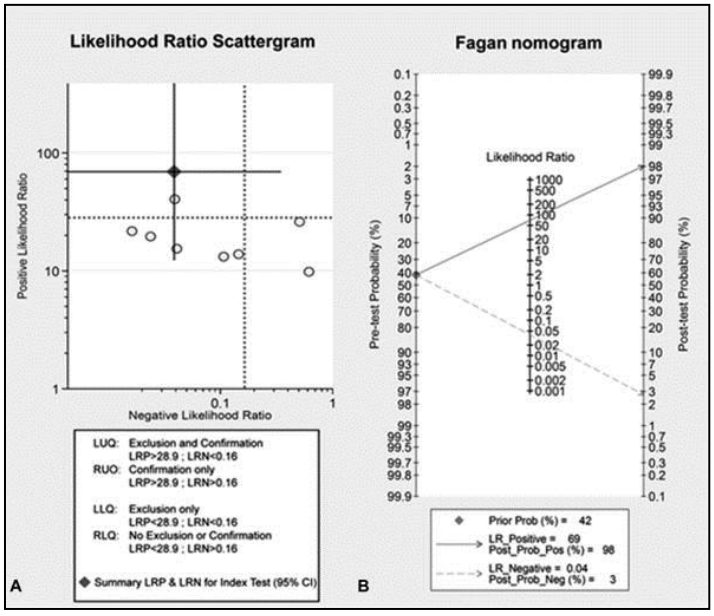

FIGURE 6 . Likelihood ratio scattergram. A, The summary point of likelihood ratio was obtained as a function of the mean sensitivity and specificity values. LRP and LRN (brown dotted straight lines) were settled at 28.9 and 0.16, respectively, splitting the graphical area into the following quadrants: 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.

Minimum Expected Conditions					Calculated Sample Size		
Sensitivity	Specificity	Overall Statistical Power	*P*	Prevalence of Disease in General Population	Subjects With Condition (n)	Subjects Without Condition (n)	Total Subjects (n)
0.80	0.98	≥0.95	<0.05	0.05	35	665	700
0.80	0.98	≥0.95	<0.05	0.10	35	315	350
0.80	0.98	≥0.90	<0.05	0.05	28	532	560
0.80	0.98	≥0.90	<0.05	0.10	28	252	280
0.80	0.98	≥0.80	<0.05	0.05	20	380	400
0.80	0.98	≥0.80	<0.05	0.10	20	180	200

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