BMJ Open

Bayesian estimation of cardiovascular autonomic neuropathy diagnostic test based on short-term heart rate variability without a gold standard

Zi-Hui Tang,[1,2] Lin Wang,[1] Fangfang Zeng,[2] Zhongtao Li,[2] Xiaoling Yu,[2] Keqin Zhang,[1] Linuo Zhou[2]

To cite: Tang Z-H, Wang L, Zeng F, *et al*. Bayesian estimation of cardiovascular autonomic neuropathy diagnostic test based on short-term heart rate variability without a gold standard. *BMJ Open* 2014;**4**: e005096. doi:10.1136/bmjopen-2014-005096

▸ Prepublication history and additional material is available. To view please visit the journal (http://dx.doi.org/10.1136/bmjopen-2014-005096).

Zi-Hui Tang, Keqin Zhang and Linuo Zhou are co-corresponding authors.

Received 20 February 2014
Revised 21 August 2014
Accepted 26 August 2014

[1]Department of Endocrinology and Metabolism, Shanghai Tongji Hospital, Tongji University School of Medicine, Shanghai, China
[2]Department of Endocrinology and Metabolism, Fudan University Huashan Hospital, Shanghai, China

Correspondence to
Dr Zi-Hui Tang;
albert.tang@163.com

ABSTRACT

Objective: To evaluate the reference values for short-term heart rate variability (HRV), estimate the performance of cardiovascular autonomic neuropathy (CAN) diagnostic tests in the absence of a gold standard, and assess CAN prevalence in our dataset.

Setting: Community and hospital health centre.

Participants: Of 2092 subjects available for data analysis, 371 healthy subjects were selected so the reference values for the short-term HRV test could be evaluated. An external dataset contained 88 subjects who completed both the short-term HRV test and Ewing's test.

Intervention: Collection of information on clinical outcome.

Primary and second outcome measures: Cardiovascular autonomic function evaluated by using the short-term HRV test and/or Ewing's test.

Results: Cut-off points of 356.13, 55.45 and 36.64 ms^2 were set for total power, low frequency and high frequency (HF), respectively. The diagnostic test for CAN based on the mentioned reference value was created. The HRV test had a high sensitivity (80.01–85.09%) and specificity (82.30–85.20%) for CAN. In addition, the non-inferiority test rejected the null hypothesis that the performance of the HRV test was inferior to that of Ewing's test ($p<0.05$). The estimated CAN prevalence was 14.92% and 29.17% in the total sample and patients with diabetes, respectively.

Conclusions: Our findings provided reference values for short-term HRV, which were used for the CAN diagnostic test with high sensitivity and specificity. The estimated CAN prevalence was high in the Chinese population.

Strengths and limitations of this study

- This was a large-scale, cross-sectional study of diagnostic tests in a Chinese population.
- This is the first study to evaluate cardiovascular autonomic neuropathy diagnosis using the short-term heart rate variability test by using Bayesian analysis without a gold standard.
- The diagnostic performance of the short-term heart rate variability test and Ewing's test was compared using a non-inferior test.
- The findings of this study can be applied to the Chinese population, but not to other ethnic groups.

INTRODUCTION

The prevalence of cardiovascular autonomic neuropathy (CAN) is increasing worldwide, particularly in the developing world.[1] The disease is a major factor contributing to the cardiovascular complications of diabetes mellitus (DM),[2] and also affects many other majority segments of the general population, such as the elderly, and patients with hypertension (HT) and metabolic syndrome (MS).[1 3 4] Individuals with previously undiagnosed CAN have an unfavourable cardiovascular risk profile, especially in terms of sudden death, indicating a higher risk of cardiovascular disease.[1]

In general, tests to assess CA function consist of the classic Ewing's test and spectral analysis of spontaneous heart rate variability (HRV).[5 6] Ewing's test includes five simple non-invasive cardiovascular autonomic reflex tests that are widely used in diabetology as well as rheumatology and gastroenterology.[7 8] Ewing's test has been reported to have high sensitivity and specificity for CAN diagnosis.[2 9] However, this test requires personnel with specialised skills and is not readily available in general practice.[9 10] Spectral analysis of HRV has the advantage of quantitatively assessing CA activity, and yields results that are similar to those produced by Ewing's test.[2 9 10] Our previous study indicated that significant negative values for the short-term HRV indices correlated with DM, HT and MS.[11] Compared with traditional methods, the short-term HRV test is simple, non-invasive and reproducible; therefore, it is easily used together with other diagnostic

tests for a large number of individuals in the general population.[2 10]

However, normal reference values for short-term HRV for assessment of CA function in the Chinese population have not been reported. There is no widely accepted gold standard approach to CAN diagnosis. Studies have been performed to evaluate the performance of new diagnostic tests using Ewing's test as a reference.[2 10 12] However, Ewing's was actually not a gold standard. Moreover, the performance of the HRV test was similar to that of Ewing's test,[2 9] so both diagnostic tests should be considered as acceptable and comparable for CA function assessment or CAN diagnosis. In addition, our previous study was performed to evaluate the performance of CAN diagnosis based on a baroreflex sensitivity test by using Bayesian analysis without a gold standard.[13] In general, the Bayesian approach to inference about a generic parameter θ combines prior information about θ with the data to obtain the posterior distribution of θ, p(θ|data). Then, one can use the mean, median or mode of this posterior distribution as an estimate of θ. Once one has obtained a sample from p(θ|data), a Monte Carlo based estimate of θ can be calculated.

This study aimed to evaluate the reference values for short-term HRV in a large cross-sectional dataset, and to estimate the sensitivities and specificities of CAN diagnostic tests using the Bayesian approach, in the absence of a gold standard, in another independent dataset. Finally, CAN prevalence was estimated in our cross-sectional dataset.

METHODS AND MATERIALS

Study population

This study is a survey of CAN factors carried out in a random sample of the middle-aged Chinese population.[11] Participants were recruited from rural and urban communities in Shanghai. Survey participants with undiagnosed CAN and aged 30–80 years were included in the study. A total of 3012 subjects were invited to a screening visit between 2011 and 2012. Some subjects were excluded from the study because of potential confounding factors that may have influenced their CA function.[11] Briefly, the exclusion criteria were as follows: (1) history or findings of arrhythmia, and hyperthyroidism or hypothyroidism; (2) pregnancy or lactation; and/or (3) serious hepatic or renal dysfunction. Complete baseline data were obtained for 2092 (69.46%) of the participants. To create an external dataset, 88 individuals with the same inclusion and exclusion criteria were recruited from another cohort, such as a healthy examination centre, to assess CA function using both the short-term HRV and Ewing's tests. Written consent was obtained from all patients before the study began. This study was approved by the Ethics Committee of Huashan Hospital, Shanghai, China.

For reference value analysis, 371 healthy subjects were selected from the dataset. The inclusion criteria included the following: (1) clinically stable condition with no previous medical history of DM, HT, dyslipidemia, coronary artery disease, cerebral stroke or heart failure; (2) fasting plasma glucose (FPG) <100 mg/dL and 2 h plasma glucose (2hPG) <140 mg/dL after a 75 g oral glucose tolerance test; (3) normal body mass index (BMI) between 18.5 and 24.9 kg/m^2; (4) triglycerides (TG) <150 mg/dL and high-density lipoprotein (HDL) cholesterol >40 mg/dL; and (5) systolic blood pressure (SBP) <140 mm Hg and diastolic blood pressure (DBP) <90 mm Hg. The exclusion criterion was the use of any medications that may affect resting HR, such as β-receptor blockers, 1 month before the study.

Measurement

Subjects were interviewed for the documentation of medical histories and medication, history of smoking habits and laboratory assessment of cardiovascular disease risk factors. All study subjects underwent a complete clinical baseline characteristics evaluation after an 8 h fast, which included: (1) history and physical examination, (2) heart rate and blood pressure (BP), (3) FPG and insulin, and (4) fasting plasma lipids. BMI was calculated as weight in kilograms divided by the square of height in metres. FPG was quantified by the glucose oxidase procedure. Serum total cholesterol, HDL cholesterol, TG levels, creatinine and uric acid were measured by an enzymatic method with a chemical analyser (Hitachi 7600-020, Tokyo, Japan). Low-density lipoprotein cholesterol levels were calculated using the Friedewald formula. At the central laboratory in our hospital, the day-to-day and inter-assay coefficients of variation for all analyses were between 1% and 3%. MS was diagnosed in individuals who met three or more of the updated National Cholesterol Education Program/Adult Treatment Panel III criteria (WHO Western Pacific Region obesity criteria).[14]

Diagnostic tests

HRV values were measured non-invasively by power spectral analysis. Subjects were studied while awake in the supine position after 20 min of rest. Testing times were from 8:00 to 11:00 in the morning. A type I FDP-1 HRV non-invasive detection system was used with V.2.0 software (Department of Biomedical Engineering of Fudan University, Shanghai, China). Electrocardiograms, respiratory signals and beat-to-beat BP were continually and simultaneously recorded for 15 min using an HMX-3C electrosphygmograph transducer (placed on the radial artery of the dominant arm) and a respiration sensor. The short-term HRV analysis was performed for all subjects using a computer-aided examination and evaluation system for spectral analysis to investigate changes in autonomic regulation. The following HRV parameters were measured by frequency domain spectral analysis[9]: total power (TP), lower frequency (LF), normalised LF (LFn), high frequency (HF) and normalised HF (HFn). The TP is the variance of the

normal-to-normal interval over a temporal segment; HF is closely related to vagal activity. The LF/HF ratio was calculated because it is considered to reflect sympathovagal balance or sympathetic modulation.[10]

Ewing's test for the detection of subclinical CAN was carried out as previously described.[5] Briefly, HRV values were analysed during three manoeuvres: deep-breathing (DB), lying-to-standing (LS) and Valsalva (V) tests. The DB test consisted of six deep respiratory cycles in 1 min. The result of the DB test was expressed as the mean value for the ratio of maximal interval between two consecutive R waves on the ECG (RR) during breathing out, over minimal RR during breathing in at each respiratory cycle. The result of the LS test was expressed as the ratio of the longest RR interval (about the 30th beat after standing up) over the shortest RR interval (about the 15th beat). The Valsalva test was performed three consecutive times, and the mean value for the Valsalva ratio was defined as the longest RR interval after Valsalva release over the shortest RR interval during the active phase of Valsalva. Cardiac parasympathetic neuropathy was considered to be present when at least one test was abnormal according to age. The other two tests investigated BP response to the LS test and to a standard handgrip test. Postural hypotension was assessed by measuring BP after 10 min in the recumbent position and again after 1 min in the standing position. Postural hypotension was defined as a drop in SBP of ≥20 mm Hg or in DBP of ≥10 mm Hg. The handgrip test consisted of determining the maximal contraction with a dynamometer and then maintaining one-third of the maximal contraction for 5 min. An increase in DBP lower than 10 mm Hg was considered to be abnormal. The three tests evaluating HRV are mainly dependent on parasympathetic control, whereas the other two tests evaluating BP response are mainly dependent on sympathetic activity. In this study, CAN was diagnosed based on at least two abnormal CA reflex test results (based on Ewing's test model or HRV test model).[2]

Statistical analysis

The Kolmogorov-Smirnov (K-S) test was used to determine whether continuous variables followed a normal distribution. Variables that were not normally distributed were log-transformed to approximate normal distribution for analysis. The results are expressed as the mean ±SD or median, unless otherwise stated. The quantiles were based on the distribution of HRV values, where the 5th, 10th and 50th percentiles were considered, and the median was the 50th quantile. Pearson and Spearman analytical methods were employed for correlation analysis of two variables. Skewed data in tables are usually reported using 2.5th and 97.5th percentiles or the median; however, in this study we have reported skewed data for HRV indices using the mean and SD because HRV parameters are often presented in this way in other studies. Additionally, we have described HRV indices using two formats in the tables. In our study, we performed correlation analysis between age and HRV having a skewed distribution using the Spearman correlation test so as not as to show log-transformed data.

We used a Bayesian latent class model to estimate the sensitivity and specificity of the HRV test and/or Ewing's test for CAN in the absence of a gold standard, as described by Branscum *et al*.[15] Latent class analysis allows characterisation of a discrete latent class (here, the true disease status) by discrete observed variables. In this model, both tests are equally considered as imperfect. There are unknown parameters about which inference must be made: the CAN population prevalence, and the sensitivity and specificity of each of the two tests. The Bayesian approach can simultaneously estimate all five unknown parameters (prevalence of CAN; sensitivity of HRV test; sensitivity of Ewings' test; specificity of HRV test; and, specificity of Ewings' test). These methods proceed in two steps: first, a prior distribution summarises the available pre-experimental information about the parameters. Subsequently, the prior distribution is updated via Bayes' theorem to a posterior distribution, using the data and the usual multinomial likelihood function. Marginal posterior densities can be derived for each parameter by integration, from which 95% marginal posterior credible intervals can be calculated. Since the integration here is analytically intractable, the Gibbs sampler, a Monte Carlo approach to calculating marginal densities, is employed. The above methods allow for simultaneous inferences to be made for all unknown parameters, which take full advantage of all the information contained in the data, as well as formally incorporated prior information, when available. See the online supplementary file for details. Data were analysed using SPSS V.16.0 and WinBUGS.14 for the Bayesian analysis. The minimum sample size estimation for this diagnostic performance analysis was 80 subjects according to the sample size estimation formula: $N=Z^2\mathrm{Sen}(1-\mathrm{Sen})/\delta^2+Z^2\mathrm{Spe}(1-\mathrm{Spe})/\delta^2$; where Z was derived from the α level (0.05 in this study), Sen (sensitivity) and Spe (specificity) were set to 0.85, respectively, and δ was set to 0.08–0.01.

Prior distributions can be estimated based on a review of the literature and/or expert opinion in the absence of data. Published evaluations of Ewing's test indicated good sensitivity (0.7–1.0) and specificity (0.7–1.0), which has a β distribution with parameters (α, β).[2 9 10 16] Previous studies demonstrated that the performance of the HRV test to assess CA activity was similar to that of Ewing's test.[10 17 18] We hypothesised that the β distribution of the sensitivity and specificity of the short-term HRV test was between 0.7 and 1.0. Finally, β of the prior distribution of prevalence was considered to be between 0.1 and 0.5.[9 10 19] The same parameters of prior distribution for the HRV test alone were estimated in the total sample, and in the DM, HT and MS patients. The two tests used here relied on the analysis of HRV attributes. As recommended by Dendukuri and Joseph,[20] the tests in the main analysis were also considered a conditionally

independent model. The particular β prior density for each test parameter was selected by matching the centre of the range with the mean of the β distribution, given by $\alpha/(\alpha+\beta)$, and matching the variance of the β distribution, given by the square root of $(\alpha\beta)/((\alpha+\beta)^2(\alpha+\beta+1))$, with one quarter of the total range.

RESULTS

The baseline characteristics of the 2092 subjects are listed in table 1. The entire sample included 905 males and 1187 females (mean age, 60.78±9.25 years). The majority of subjects had never smoked (85.37%), and the prevalence of HT, DM and MS was 46.65%, 21.33% and 39.82%, respectively, in the entire sample. A total of 371 healthy subjects, consisting of 78 males and 293 females, were selected for reference value analysis. The mean age of the healthy subjects (56.5±8.75 years) was younger than that of the entire sample. The demographic parameters, blood glucose parameters, lipid profiles and medical histories of the healthy subjects were better than those of the entire sample. The HRV indices of healthy subjects were significantly higher than those of the entire sample. The mean age of external subjects was younger than that of the entire sample. However, the other demographics parameters, glucose parameters, lipid profiles, HRV indices and medical histories were similar to those of the entire sample.

Reference values for short-term HRV

No normal distribution results were found in HRV indices using K-S tests ($p<0.05$ for all, data not shown). In this study, we set the 5th percentile as the cut-off point for TP, LF and HF indices. The normal value of LF/HF ranged from the 2.5th to the 97.5th percentile. Age had a strong negative correlation with HRV (figure 1). TP, LF and HF had significant negative correlations with age ($r=-0.111$–0.291, $p<0.05$ for all). No significant correlation between age and LF/HF was found ($p>0.05$). Reference values for the total sample and subjects stratified by age were calculated and are listed in table 2. In the total sample, the reference value for TP was more than 356.13 ms^2. Cut-off points of 55.45 and 36.64 ms^2 were set for LF and HF, respectively. The cut-off points for LFn and HFn were 6.40 and 4.83 nu, respectively. The reference values for LF/HF ranged

Table 1 Baseline characteristics of subjects

Variables	Entire sample	Healthy subjects	External dataset
Demographic information			
N	2092	371	88
Age, years	60.42±8.68	56.5±8.75	56.61±9.26
Male gender, %	705 (33.7%)	78 (21.02%)	38 (43.18%)
BMI, kg/m^2	24.21±3.36	21.57±1.99	23.81±2.72
WC, cm	85.07±9.70	77.08±6.83	85.09±7.12
SBP, mm Hg	127.62±18.68	114.6±10.93	117.85±14.62
DBP, mm Hg	79.83±9.69	73.81±6.92	79.23±9.59
Laboratory assays			
FPG, mmol/L	5.53±1.81	4.64±0.59	6.84±2.38
PBG, mmol/L	7.67±3.56	5.26±1.11	10.29±4.51
FINS, IU/L	7.19±11.82	32.72±13.67	8.77±19.31
TC, mmol/L	5.32±1	5.19±0.96	5.35±1.05
TG, mmol/L	1.71±0.98	1.1±0.31	1.86±1.07
HDL, mmol/L	1.36±0.32	1.55±0.33	1.30±0.30
LDL mmol/L	3.19±0.77	3.04±0.77	3.14±0.79
UA, μmol/L	281.21±83.79	240.52±68.77	292.34±82.65
HRV measurement			
HR, bpm	72.42±10.13	68.39±8.61	72.23±9.80
TP, ms^2	873.95±702.47	1127.33±697.28	883.69±935.65
LF, ms^2	190.98±207.88	241.35±204.92	164.04±220.97
LFn, nu	21.33±10.66	21.44±10.56	17.13±8.82
HF, ms^2	183.05±219.43	245.93±230.26	208.99±399.68
HFn, nu	20.67±13.25	21.85±12.83	20.84±16.23
LF/HF	1.70±1.98	1.53±1.79	1.50±1.64
Medical history			
Smoking yes, %	306 (14.63%)	35 (9.43%)	13 (14.77%)
HT yes, %	976 (46.65%)	0 (0%)	32 (36.36%)
DM yes, %	446 (21.33%)	0 (0%)	53 (60.23%)
MS yes, %	833 (39.82%)	0 (0%)	34 (38.64%)

BMI, body mass index; DBP, diastolic blood pressure; DM, diabetes; FINS, fasting blood insulin; FPG, fasting plasma glucose; HDL, high-density lipoprotein cholesterol; HF, high frequency; HFn, normalised HF; HR, heart rate; HRV, heart rate variability; HT, hypertension; LDL, low density lipoprotein cholesterol; LF, low frequency; LFn, normalised LF; MS, metabolic syndrome; PBG, plasma blood glucose; SBP, systolic blood pressure; TC, serum total cholesterol; TG, triglyceride; TP, total power of variance; UA, uric acid; WC, waist circumference.

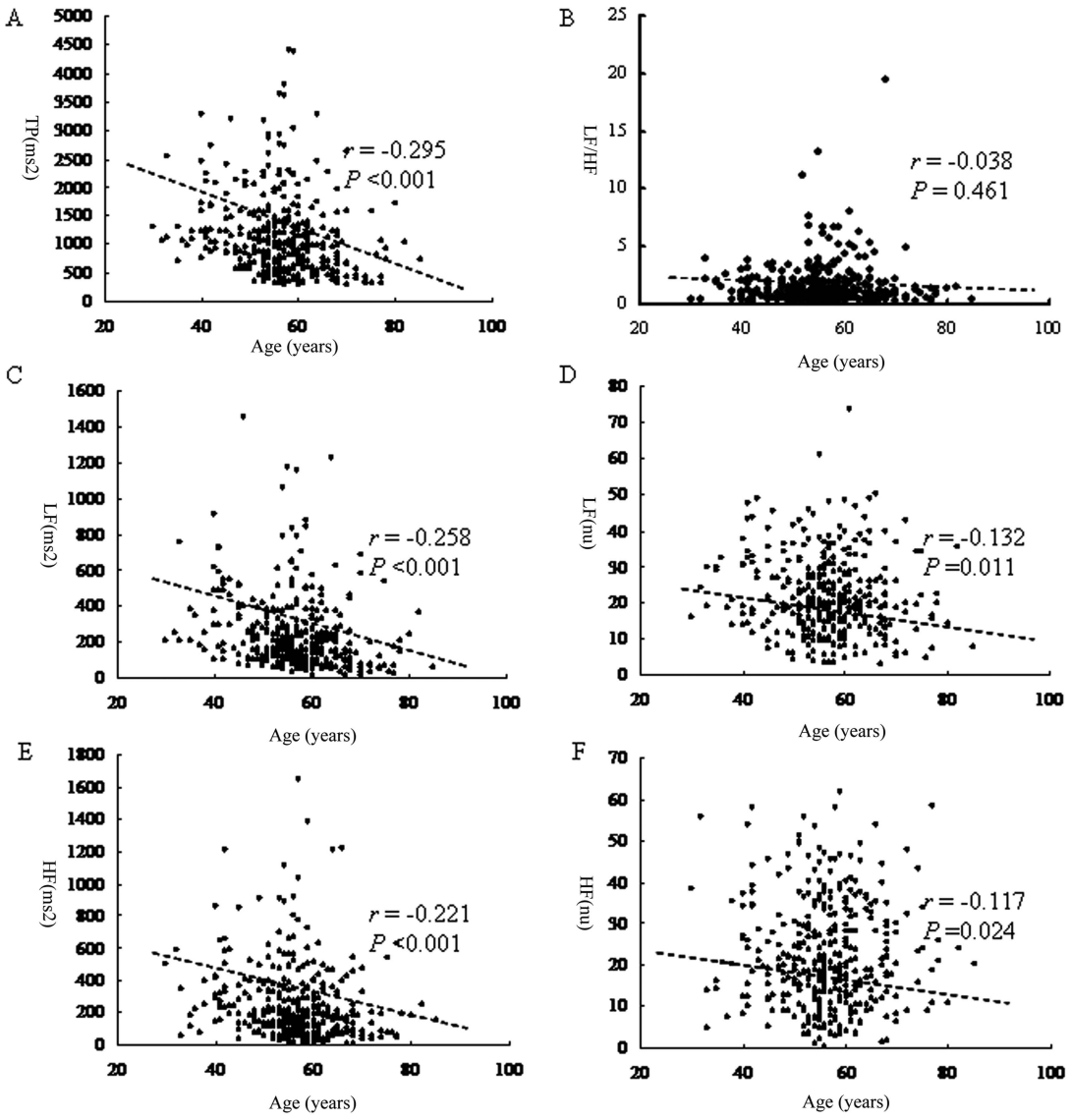

Figure 1 Results of correlation analysis between age and parameters of short-term heart rate variability. (A) Correlation analysis between age and TP (r=−0.295 and p<0.001); (B) correlation analysis between age and LF/HF (r=−0.038 and p=0.461); (C) correlation analysis between age and LF (ms^2) (r=−0.258 and p<0.001); (D) correlation analysis between age and LF (nu) (r=−0.132 and p=0.011); (E) correlation analysis between age and HF (ms^2) (r=−0.221 and p<0.001); and (F) correlation analysis between age and HF (nu) (r=−0.117 and p=0.024). HF, high frequency; LF, low frequency; TP, total power.

from 0.3 to 6.5. In our study, the HRV test (model 1) was based on the uniform reference values for HRV indices used to set the diagnostic criteria for CAN, while the HRV test (model 2) was based on reference values stratified by age.

Bayesian estimation of diagnostic tests

Of the 88 subjects in the external dataset, 31 and 43 were diagnosed with CAN using the HRV test (model 1) alone and Ewing's test alone, respectively (table 3). Using both tests, 21 subjects were diagnosed with CAN, while 35 subjects were diagnosed as being free of CAN. Using the HRV test (model 2), 33 subjects were diagnosed with CAN. The prior parameters are listed in table 4.

Median posterior CAN prevalence using the HRV test (model 1) alone and Ewing's test alone was estimated as 29.06% and 38.55%, respectively. The median posterior sensitivity and specificity of the HRV test (model 1) alone were 85.09% and 85.20%, respectively, while they were 87.13% and 80.46%, respectively, for Ewing's test alone. When both tests were combined with the conditional independence model, the median posterior CAN prevalence was 32.15% (95% CI 19.36% to 46.92%), and the median posterior sensitivity and specificity of the HRV test (model 1) were 81.13% (95% CI 62.9% to 93.97%) and 85.17% (95% CI 74.36% to 94.03%), respectively. Ewing's test had an apparently higher sensitivity (85.53%) but a lower specificity (73.55%) compared with the HRV test. When both tests were combined with the conditional dependence model, the

Table 2 Reference values for short-term HRV

Variable	Total sample	Aged ≤45 years	Aged 46–55 years	Aged 55–65 years	Aged >65 years
N	371	42	120	160	49
TP, ms^2					
Mean±SD	1127.33±697.28	1425.87±602.81	1163.64±587.5	1112.14±801.26	832.07±532.91
5th percentile	356.13	786.34	463.91	345.92	317.71
10th percentile	441.58	850.57	568.35	419.09	326.79
50th percentile	972.82	1225.39	1058.10	902.77	671.60
LF, ms^2					
Mean±SD	241.35±204.92	370.77±207.51	233.69±200.62	221.6±207.04	164.69±155.73
5th percentile	55.45	84.34	70.92	55.11	20.82
10th percentile	68.75	113.90	88.18	64.78	34.52
50th percentile	173.44	345.18	181.56	167.22	103.24
LFn, nu					
Mean±SD	21.44±10.56	25.69±10.61	23.48±9.92	21.84±10.71	18.85±10.66
5th percentile	6.40	9.93	6.68	6.20	4.86
10th percentile	9.65	13.43	10.49	9.82	6.05
50th percentile	19.51	25.33	19.95	19.50	17.46
HF, ms^2					
Mean±SD	245.93±230.26	348.93±241.48	246.33±203.98	232.74±245.16	199.74±211.14
5th percentile	36.64	57.37	38.64	32.09	27.07
10th percentile	53.87	105.26	59.18	53.96	48.57
50th percentile	183.97	283.00	194.54	156.30	147.58
HFn, nu					
Mean±SD	21.85±12.83	24.79±13.79	22.07±12.7	21.37±12.66	20.82±12.87
5th percentile	4.83	5.48	5.05	4.78	4.15
10th percentile	7.04	8.75	7.10	6.23	6.22
50th percentile	19.09	20.82	18.73	18.64	18.35
LF/HF					
Mean±SD	1.53±1.79				
2.5th percentile	0.30				
50th percentile	1.00				
97.5th percentile	6.50				

HF, high frequency; HFn, normalised HF; HRV, heart rate variability; LF, low frequency; LFn, normalised LF; TP, total power.

median posterior CAN prevalence was 30.60% (95% CI 15.26% to 50.65%) and the median posterior sensitivity and specificity of the HRV test were 80.01% (95% CI 54.68% to 94.64%) and 82.30% (95% CI 69.34% to 93.63%), respectively. There were modest correlations between the HRV test and Ewing's test (ρ_P=0.317 and ρ_N=0.319; ρ_P = correlation coefficients of the sensitivities of two tests; ρ_N = correlation coefficients of the specificities of two tests).

Table 3 Results of short-term heart rate variability (HRV) test and Ewing's test for cardiovascular autonomic neuropathy in 88 subjects

Model		Ewing's test +	Ewing's test −	Total
HRV test (model 1)	+	21	10	31
	−	22	35	57
Total		43	45	88
HRV test (model 2)	+	23	10	33
	−	20	35	55
Total		43	45	88

Similar parameters were found for the HRV test (model 2) for CAN (table 4). Generally, the median posterior sensitivities and specificities of the HRV test were over 80% in all models. Higher sensitivities and lower specificities for Ewing's test were found in all models, compared with those of the HRV test. The posterior Youden indices of the HRV test were higher than those of Ewing's test in all models. In combined tests, we compared the parameters (mean sensitivity and mean specificity) of performance of both diagnostic tests by using a non-inferiority test that rejected the hypothesis that the performance of the HRV test was inferior to that of Ewing's test ($p<0.05$ for all parameters in two models, table 5).

Estimated CAN prevalence in different groups

In the entire sample, 387 and 465 subjects were diagnosed with CAN using the HRV test (model 1) and the HRV test (model 2) alone, respectively (table 6). The median posterior sensitivities and specificities of the HRV test (model 1) alone for CAN were high in four different groups (sensitivities >80% and specificities >85% for all). The median posterior CAN prevalence in the total sample was estimated at 14.92%. The estimated median

Table 4 Marginal prior and posterior medians and lower and upper limits of the posterior equal tailed 95% credible interval for parameters of diagnostic tests

Variables	Prior information	HRV test alone	Ewing's test alone	Both tests combined (independence)		Both tests combined (dependence)	
				HRV test	Ewing's test	HRV test	Ewing's test
HRV test (model 1)							
Prevalence, %	30.00 (12.00–48.00)	29.02 (15.23–44.59)	38.64 (21.35–54.35)	32.15 (19.36–46.92)		30.60 (15.26–50.65)	
Sensitivity, %	85.00 (72.00–98.00)	85.09 (66.85–95.42)	87.13 (72.26–96.11)	81.13 (62.9–93.97)	85.53 (71.31–95.33)	80.01 (54.68–94.64)	80.82(62.86–94.25)
Specificity, %	85.00 (72.00–98.00)	85.20 (71.7–95.14)	80.46 (63.84–93.57)	85.17 (74.36–94.03)	73.55 (61.19–86.46)	82.30 (69.34–93.63)	69.13(55.64–89.28)
Youden index, %	70.00 (44.00–96.00)	70.29 (38.55–90.56)	67.59 (36.10–89.68)	66.3 (37.26–88)	59.08 (32.5–81.79)	62.24 (23.02–88.27)	49.95(18.5–83.53)
PPV, %	70.83 (25.96–97.84)	70.2 (29.79–94.05)	73.74 (35.17–94.68)	72.16 (37.07–93.29)	60.51 (30.61–86.16)	66.57 (23.97–93.85)	53.58(20.33–90.02)
NPV, %	92.97 (94.96–98.15)	93.31 (92.33–96.27)	90.85 (89.45–95.28)	90.5 (89.3–94.64)	91.47 (89.88–95.44)	90.3 (89.26–94.45)	89.1(89.27–93.8)
ρ_P						0.317 (0.016–0.838)	
ρ_N						0.319 (0.014–0.806)	
HRV test (model 2)							
Prevalence, %	30.00 (12.00–48.00)	30.84 (16.32–45.80)	38.64 (21.35–54.35)	33.88 (21.21–48.76)		32.32 (17.10–51.11)	
Sensitivity, %	85.00 (72.00–98.00)	85.44 (68.67–95.58)	87.13 (72.26–96.11)	82.17 (65.06–94.46)	85.92 (71.82–95.54)	80.34 (56.63–94.57)	81.02(63.43–94.20)
Specificity, %	85.00 (72.00–98.00)	84.65 (70.93–95.00)	80.46 (63.84–93.57)	85.01 (73.90–94.13)	75.23 (62.83–88.22)	81.75 (68.21–93.36)	70.60(56.90–89.15)
Youden index, %	70.00 (44.00–96.00)	70.09 (39.60–90.58)	67.59 (36.10–89.68)	67.18 (38.96–88.59)	61.15 (34.65–83.76)	61.89 (24.84–87.93)	51.62(20.33–83.35)
PPV, %	70.83 (25.96–97.84)	71.28 (31.54–94.17)	73.74 (35.17–94.68)	73.75 (40.16–93.87)	63.99 (34.22–88.53)	67.71 (26.87–93.71)	56.82(23.29–90.08)
NPV, %	92.97 (94.96–98.15)	92.88 (92.07–96.22)	90.85 (89.45–95.28)	90.30 (88.71–94.70)	91.25 (89.23–95.41)	89.60 (88.41–94.27)	88.62(88.29–93.63)
ρ_P						0.332 (0.017–0.820)	
ρ_N						0.329 (0.016–0.835)	

The HRV test (model 1) was based on uniform reference values, while the HRV test (model 2) was based on reference values stratified by age.
HRV, heart rate variability; NPV, negative predictive value; PPV, positive predictive value; ρ_P, correlation coefficients of the sensitivities of two tests; ρ_N, correlation coefficients of the specificities of two tests.

posterior CAN prevalence values were 29.17%, 20.04% and 21.16% in DM, HT and MS patients, respectively. The estimated CAN prevalences based on the HRV test (model 2) in different groups are listed table 6.

DISCUSSION

A large-scale, population-based, cross-sectional study was conducted to evaluate the reference values for the short-term HRV test and the Bayesian estimate of the performance of diagnostic tests for CAN among 2092 participants in the Chinese population. This sample was an adequate representation of the Chinese population, and the reference values may work similarly well outside the areas studied in China.[21 22] Importantly, we first carried out a performance analysis of the short-term HRV test for CAN by using Bayesian approaches in the general Chinese population. It is crucial to understand that the HRV test and Ewing's test are similar diagnostic tests. In addition, evaluation of the performance of the short-term HRV test would be inappropriate if Ewing's test was used as the reference standard. In the absence of a gold standard, the Bayesian approach can be applied to estimate diagnostic tests.

Reference value analysis

The results of this study most likely reflect typical HRV patterns for healthy subjects. The HRV parameters provide general information on CA function. Cut-off points for the HRV indices for the total sample are reported (table 2). Bigger *et al*[23] reported that in a Caucasian sample, the reference values were higher than in our study. Recently, Kim and Woo[24] conducted a study to examine the normal reference values for short-term HRV measurements in a large Korean cohort (>3000 healthy participants). In the total sample, the 10th percentiles of TP, LF and HF were 347, 74.5 and 38.2 ms^2, respectively. The values for the 10th to 90th percentiles of LF/HF ranged from 0.6 to 5.1, suggesting that our findings are consistent with these results. Establishment of normal reference values may therefore provide important evidence for clinical evaluation of CAN. In this study, evaluation of the relationship between HRV parameters and age using correlation analysis indicated that HRV indices were independent of gender but decreased with age. Several previous studies suggested that age should be considered as an independent determinant for HRV.[25 26] Voss *et al*[25] determined that the reference values were for short-term HRV analysis in 2000 individuals, showing 216 and 94 ms^2 for TP in the younger group (<50 years) and older group (≥50 years), respectively. The younger group had lower quartile values for LF and HF of 62.88 and 30.75 ms^2, respectively. Kim and Woo[24] reported reference values stratified by age that were similar to those in our study. Our findings were good representations, and we recommend cut-off points for the normal reference values of HRV parameters.

Table 5 Comparison of performance of the HRV test and Ewing's test in both tests combined

	Both tests combined (independence)					Both tests combined (dependence)				
	HRV test		Ewing's test			HRV test		Ewing's test		
Parameter	Mean	SE	Mean	SE	p Value	Mean	SE	Mean	SE	p Value
HRV test (model 1)										
Sensitivity	0.804	0.081	0.853	0.063	0.019	0.782	0.108	0.802	0.080	0.048
Specificity	0.849	0.051	0.753	0.064	<0.001	0.821	0.062	0.700	0.085	<0.001
HRV test (model 2)										
Sensitivity	0.815	0.077	0.853	0.062	0.011	0.787	0.103	0.804	0.080	0.023
Specificity	0.848	0.052	0.753	0.064	<0.001	0.815	0.065	0.712	0.081	<0.001

A non-inferiority test tested the hypothesis that the performance of the HRV test was inferior to Ewing's test; $\delta>0$ is the non-inferiority margin of clinical interest, which is set as a quarter of the value of the parameter of Ewing's test in this study.
HRV, heart rate variability.

Bayesian estimation of diagnostic tests

An interesting finding is that the diagnostic test for CAN based on short-term HRV showed high sensitivity and specificity. In the external dataset, the estimated median sensitivities and specificities of the HRV test for CAN were over 80% in all models. In the total sample, the estimated median sensitivity of the HRV model alone was greater than 83%. Similar results were found in DM, HT and MS patients. Moreover, the estimated median specificities of the HRV test were over 85% in four different subgroups. These findings support the fact that the HRV test is an efficient tool for the diagnosis of CAN in individuals with metabolic disorders. It was very important to evaluate the performance of the HRV test in both tests combined with the independence and dependence models using the Bayesian approach. The combined tests allowed for sharper inferences to be drawn.[27] Studies were performed to evaluate the performance of new diagnostic tests using Ewing's as a reference. However, Ewing's is a non-invasive test for CAN and is actually not a gold standard. Using the Bayesian approach, in the absence of a gold standard, simultaneous inferences about the performance of each diagnostic test are possible. Additionally, Bayesian estimation of the parameters of a diagnostic test needs prior information. This is more suitable for clinical research because it is easier for experts with relevant clinical experience to estimate prior parameters accurately. In this study, precise and accurate posterior parameters should be derived from appropriate prior sensitivity of HRV test parameters. A non-inferiority test indicated that the sensitivity of HRV test parameter was not inferior to that of Ewing's test. The Youden index of the HRV test was higher than that of Ewing's test. These findings demonstrate that the performance of the HRV test was not inferior to Ewing's test. To our knowledge, this is the first study to report that in the absence of a gold standard, the short-term HRV test for CAN diagnosis had high sensitivity and specificity and was not inferior to Ewing's test using Bayesian analysis in a general Chinese population. This finding is important for the clinical diagnosis of CAN in the general population. We recommend the HRV test, based on the uniform normal reference values (model 1), as an acceptable diagnostic test for CAN.

Estimation of CAN prevalence

In our study sample, when the HRV test (model 1) was used alone, the prevalence of CAN was estimated to be 14.92% in the general population. In patients with DM, its prevalence was estimated to be 29.17%. The estimated

Table 6 Estimated prevalence of cardiovascular autonomic neuropathy (CAN) in different groups

			Prevalence (%)		Sensitivity† (%)		Specificity† (%)	
Group	CAN*	Population	Median	95% CI	Median	95% CI	Median	95% CI
HRV test (model 1)								
Total sample	387	2092	14.92	9.63 to 26.69	82.76	61.87 to 94.74	92.32	86.56 to 97.42
DM	149	446	29.17	18.59 to 43.48	84.33	67.41 to 95.30	86.57	74.64 to 95.20
HT	241	976	20.04	11.78 to 32.22	83.42	65.05 to 94.67	89.82	80.94 to 96.45
MS	204	833	21.16	12.78 to 34.69	83.65	65.86 to 95.15	90.84	82.48 to 96.63
HRV test (model 2)								
Total sample	465	2092	18.26	9.63 to 26.69	83.12	63.70 to 94.84	90.99	83.98 to 97.00
DM	162	446	31.60	18.59 to 43.48	84.83	67.57 to 95.46	85.96	74.64 to 95.34
HT	263	976	22.35	11.78 to 32.22	83.56	65.12 to 95.00	88.86	80.02 to 96.28
MS	241	833	23.94	12.78 to 34.69	84.02	66.17 to 95.12	87.99	78.48 to 96.15

*Subjects diagnosed with CAN using the HRV test.
†Bayesian estimation of the HRV test alone model in different groups.
DM, patients with diabetes mellitus; HRV, heart rate variability; HT, patients with hypertension; MS, patients with metabolic syndrome.

CAN prevalence in patients with DM was found to be 20–50% in previous reports,[2 9] indicating that our result was consistent with these studies. In hypertensive individuals, CAN prevalence was estimated to be 20.04%. Our previous studies demonstrated that BP and HT were strongly associated with a low HRV.[22] Laitinen *et al*[28] reported that the prevalence of parasympathetic dysfunction was 25% in subjects with central obesity and in persons with impaired glucose tolerance. In our study, the estimated CAN prevalence was 21.16% in the MS population. Our findings supported evidence indicating that CAN has become a serious public problem in China. A higher prevalence of this disease was found in special subgroups.

Several limitations of this study warrant comment. This study does not cover age groups below or above 30–90 years of age. Additionally, a cross-sectional study for the determination of normal reference values requires a larger sample size and a wider geographical spread. Furthermore, the normal reference values of short-term HRV established in this study need to be verified in future follow-up studies. Finally, it is important to mention that our study was performed in Chinese individuals, and our findings may not be relevant to people of other ethnicities.

In conclusion, this study provided reference values for short-term HRV that were applied to the CAN diagnostic test with high sensitivity and specificity. Moreover, our findings offered evidence that the HRV test was not inferior to the traditional Ewing's test for CAN. The estimated CAN prevalence was high in the general Chinese population, and more frequent in individuals with DM, HT and MS. CAN is now a major public health problem in China, and strategies to prevent and treat it are required.

Contributors Z-HT designed the study, analysed the data and wrote the manuscript. LW, ZL, FZ and XY contributed reagents, materials and/or analysis tools. KZ and LZ conceived and designed the study.

Funding This study was supported by a grant from China National Grant on Science and Technology (grant number: 30570740).

Competing interests None.

Ethics approval This study was approved by the Ethics Committee of Huashan Hospital, Shanghai, China.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

1. Hazari MA, Khan RT, Reddy BR, *et al*. Cardiovascular autonomic dysfunction in type 2 diabetes mellitus and essential hypertension in a South Indian population. *Neurosciences (Riyadh)* 2012;17:173–5.
2. Spallone V, Ziegler D, Freeman R, *et al*; on behalf of the Toronto Consensus Panel on Diabetic Neuropathy. Cardiovascular autonomic neuropathy in diabetes: clinical impact, assessment, diagnosis, and management. *Diabetes Metab Res Rev* 2011. doi:10.1002/dmrr.1239 [Epub ahead of print 22 Jun 2011].
3. Garruti G, Giampetruzzi F, Vita MG, *et al*. Links between metabolic syndrome and cardiovascular autonomic dysfunction. *Exp Diabetes Res* 2012;2012:615835.
4. Iodice V, Low DA, Vichayanrat E, *et al*. Cardiovascular autonomic dysfunction in MSA and Parkinson's disease: similarities and differences. *J Neurol Sci* 2011;310:133–8.
5. Ewing DJ, Martyn CN, Young RJ, *et al*. The value of cardiovascular autonomic function tests: 10 years experience in diabetes. *Diabetes Care* 1985;8:491–8.
6. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. *Eur Heart J* 1996;17:354–81.
7. Smith SA. Reduced sinus arrhythmia in diabetic autonomic neuropathy: diagnostic value of an age-related normal range. *Br Med J (Clin Res Ed)* 1982;285:1599–601.
8. MacGilchrist AJ, Reid JL. Impairment of autonomic reflexes in cirrhosis. *Am J Gastroenterol* 1990;85:288–92.
9. Pop-Busui R. Cardiac autonomic neuropathy in diabetes: a clinical perspective. *Diabetes Care* 2010;33:434–41.
10. Vinik AI, Ziegler D. Diabetic cardiovascular autonomic neuropathy. *Circulation* 2007;115:387–97.
11. Li Z, Tang ZH, Zeng F, *et al*. Associations between the severity of metabolic syndrome and cardiovascular autonomic function in a Chinese population. *J Endocrinol Invest* 2013;36:993–9.
12. Maser RE, Lenhard MJ. Cardiovascular autonomic neuropathy due to diabetes mellitus: clinical manifestations, consequences, and treatment. *J Clin Endocrinol Metab* 2005;90:5896–903.
13. Tang ZH, Zeng F, Yu X, *et al*. Bayesian estimation of cardiovascular autonomic neuropathy diagnostic test based on baroreflex sensitivity in the absence of a gold standard. *Int J Cardiol* 2014;171:e78–80.
14. Grundy SM, Hansen B, Smith SC Jr, *et al*. Clinical management of metabolic syndrome: report of the American Heart Association/National Heart, Lung, and Blood Institute/American Diabetes Association conference on scientific issues related to management. *Circulation* 2004;109:551–6.
15. Branscum AJ, Gardner IA, Johnson WO. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Prev Vet Med* 2005;68:145–63.
16. Straub RH, Lang B, Palitzsch KD, *et al*. Estimation of the cut-off value in cardiovascular autonomic nervous function tests: not-age-related criteria or the age-related 5th percentile. *J Diabetes Complications* 1997;11:145–50.
17. Neil HA, Thompson AV, John S, *et al*. Diabetic autonomic neuropathy: the prevalence of impaired heart rate variability in a geographically defined population. *Diabet Med* 1989;6:20–4.
18. Risk M, Bril V, Broadbridge C, *et al*. Heart rate variability measurement in diabetic neuropathy: review of methods. *Diabetes Technol Ther* 2001;3:63–76.
19. Ziegler D, Dannehl K, Muhlen H, *et al*. Prevalence of cardiovascular autonomic dysfunction assessed by spectral analysis, vector analysis, and standard tests of heart rate variation and blood pressure responses at various stages of diabetic neuropathy. *Diabet Med* 1992;9:806–14.
20. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 2001;57:158–67.
21. Bian RW, Lou QL, Gu LB, *et al*. Delayed gastric emptying is related to cardiovascular autonomic neuropathy in Chinese patients with type 2 diabetes. *Acta Gastroenterol Belg* 2011;74:28–33.
22. Ding W, Zhou L, Bao Y, *et al*. Autonomic nervous function and baroreflex sensitivity in hypertensive diabetic patients. *Acta Cardiol* 2011;66:465–70.
23. Bigger JT Jr, Fleiss JL, Steinman RC, *et al*. RR variability in healthy, middle-aged persons compared with patients with chronic coronary heart disease or recent acute myocardial infarction. *Circulation* 1995;91:1936–43.
24. Kim GM, Woo JM. Determinants for heart rate variability in a normal Korean population. *J Korean Med Sci* 2011;26:1293–8.
25. Voss A, Heitmann A, Schroeder R, *et al*. Short-term heart rate variability—age dependence in healthy subjects. *Physiol Meas* 2012;33:1289–311.
26. Zhang J. Effect of age and sex on heart rate variability in healthy subjects. *J Manipulative Physiol Ther* 2007;30:374–9.
27. Joseph L, Gyorkos TW, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am J Epidemiol* 1995;141:263–72.
28. Laitinen T, Lindstrom J, Eriksson J, *et al*. Cardiovascular autonomic dysfunction is associated with central obesity in persons with impaired glucose tolerance. *Diabet Med* 2011;28:699–704.

Supplementary Appendix

This document has been provided by the authors to give reviewers additional information about their work.

Supplement to: Tang ZH, Wang L, Zeng F, Li Z, Yu X, Zhang K, Zhou L. Bayesian Estimation of Cardiovascular Autonomic Neuropathy Diagnostic Test based on short-term Heart Rate Variability without a Gold Standard

Outline of Contents

1. Overview

W We used a Bayesian latent class model to estimate the sensitivity and specificity of HRV test or/and Ewing's test for CAN in the absence of a gold standard, as described by Branscum et al. [1]. The latent class analysis allows the characterization of a discrete latent class (here the true disease status) by discrete observed variables. In this model, both tests are equally considered as imperfect. There are unknown parameters about which inference must be made: the CAN population prevalence, and the sensitivity and specificity of each of the two tests. Bayesian approach can simultaneously estimate all five unknown parameters. These methods proceeds in two steps: first, a prior distribution summarizes the available pre-experimental information about the parameters. Subsequently, the prior distribution is updated via Bayes Theorem to a posterior distribution, using the data and the usual multinomial likelihood function. Marginal posterior densities can be derived for each parameter by integration, from which 95% marginal posterior credible intervals can be calculated. Since the integration here is analytically intractable, the Gibbs Sampler, a Monte Carlo approach to calculating marginal densities, is employed. The above methods allow for simultaneous inferences to be made for all unknown parameters, which takes full advantage of all the information contained in the data, as well as formally incorporating prior information, when available. Data were analyzed using SPSS16.0 (USA) and WinBUGS 14 for the Bayesian analysis.

2. Prior distribution

Prior distributions can be estimated based on a review of the literature and/or expert opinion in the absence of data. Published evaluations of the Ewing's test indicated a good sensitivity (0.7 to 1.0) and specificity (0.7 to 1.0), which has a beta distribution with parameters (α, β) [2,3,4,5]. Previous studies demonstrated that performances of HRV to assess CA activity are similar to those of Ewing's test [4,6,7]. We made a hypothesis for the short-term HRV test with sensitivity and specificity of a beta distribution between 0.7 and 1.0, respectively. Finally, the prior distribution of prevalence was considered beta between 0.1 and 0.5 [3,4,8]. The same parameters of prior distribution for HRV test alone were estimated in total sample, DM, HT and MS patients. The two tests used here rely on analysis of HRV attributes. As recommended by Dendukuri et al. [9], in the main analysis the tests were also considered conditionally independent model. The particular beta prior density for each test parameter was selected by matching the center of the range with the mean of the beta distribution, given by $\alpha/(\alpha+\beta)$, and matching the variance of the beta distribution, given by square root of $(\alpha\beta)/((\alpha+\beta)^2(\alpha+\beta+1))$ with one quarter of the total range.

3. One diagnostic test

Let A and B be the observed number of positive and negative test results, respectively, in the sample of A + B = N subjects (Table a1). Let Y_1 and Y_2 be the information that is missing when there is no gold standard, that is, the number of true positive test results out of A and B, respectively. Thus, Y_1 is the number of true positives, and Y_2 is the number of false negatives. Such missing information has been termed "latent data".

Table a1: Observed and latent data in the case of one diagnostic test In the absence of a gold standard, presented in a 2 x 2 table

		True		Total
		Positive	Negative	
Test	Positive	Y_1	A-Y_1	A
	Negative	Y_2	B-Y_2	B
	Total	Y_1+Y_2	N-Y_1-Y_2	N

The likelihood function of the observed and latent data is given by

$$L(A, B, Y_1, Y_2 | \pi, Sen, Spe) = [\pi Sen]^{Y_1}[\pi(1-Sen)]^{Y_2}[(1-\pi)(1-Spe)]^{A-Y_1}[(1-\pi)Spe]^{B-Y_2}.$$

Prior information in the form of a beta density will be assumed. A random variable ($0 \le \theta \le 1$) has a beta distribution with parameters (α, β) if it has a probability density given by

$$f(\theta;\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$, where $B(\alpha, \beta)$, the beta function evaluated at (α, β), is the normalizing constant. This family of distributions was selected since its region of positive density, from 0 to 1, matches the range of all parameters of interest in this study. In addition, it also has the advantage of being the conjugate prior distribution for the binomial likelihood, a property that simplifies the derivation of the posterior distributions. Let (α_π, β_π), (α_{Sen}, β_{Sen}), and (α_{Spe}, β_{Spe}) represent the prior beta parameters for π, Sen and Spe, respectively. Since the joint posterior distribution is proportional to the product of the likelihood function and the prior distribution.

Inference is possible using a Gibbs sampler algorithm. The basic idea is as follows.

Conditional on knowing the exact values of the prevalence and all diagnostic test parameters, it is possible to derive posterior distributions of the latent data Y_1 and Y_2. Conversely, if Y_1 and Y_2 are known, then deriving posterior distributions of the prevalence and diagnostic test parameters given the prior distributions requires only a straightforward application of Bayes' theorem. An algorithm that alternates between these two steps can thus be devised, similar in spirit to the expectation maximization algorithm that is commonly used in latent class analysis. The Gibbs sampler algorithm provides random samples from the marginal posterior densities of each parameter of interest. These random samples can then be used to reconstruct the marginal posterior densities, or summaries of these densities, such as their means, medians, or standard deviations, as well as probability interval summaries.

Implementation of the Gibbs sampler requires the specification of the full conditional distributions of the parameters, i.e., the conditional distributions of each parameter given the values of all of the other parameters. It is straightforward to show from likelihood function that the following conditional distributions must hold:

$$Y_1 | A, \pi, Sen, Spe \sim Binomial(A, \frac{\pi Sen}{\pi Sen + (1-\pi)(1-Spe)}) \quad \text{(app1.1)}$$

$$Y_2 | B, \pi, Sen, Spe \sim Binomial(B, \frac{\pi(1-Sen)}{\pi(1-Sen) + (1-\pi)Spe}) \quad \text{(app1.2)}$$

$$\pi | A, B, Y_1, Y_2, \alpha_\pi, \beta_\pi \sim Beta(Y_1 + Y_2 + \alpha_\pi, A + B - Y_1 - Y_2 + \beta_\pi) \quad \text{(app 1.3)}$$

$$Sen | Y_1, Y_2, \alpha_{Sen}, \beta_{Sen} \sim Beta(Y_1 + \alpha_{Sen}, Y_2 + \beta_{Sen}) \quad \text{(app 1.4)}$$

and

$$Spe | A, B, Y_1, Y_2, \alpha_{Spe}, \beta_{Spe} \sim Beta(B - Y_2 + \alpha_{Spe}, A - Y_1 + \beta_{Spe}) \quad \text{(app 1.5)}$$

The Gibbs sampler operates as follows. Arbitrary starting values are chosen for each parameter. A sample of size m is then drawn from each full conditional distribution, in turn. The sampled values from the previous iterations are used in the conditional distributions for subsequent iterations. A cycle of the algorithm is completed when all conditional distributions have been sampled at least once. The entire cycle is repeated a large number of times. The random samples thus generated for each parameter can be regarded as a random sample from the correct posterior marginal distribution. For the above model, Y_1 and Y_2 are

generated from expressions app 1.1 and app1.2, respectively, given the starting values of the other parameters. Then, π is generated from equation app1.3 conditional on the Y_1 and Y_2 variates just sampled. Drawing Sen and Spe from densities given in expressions app1.4 and app1.5, respectively, using the same values of Y_1 and Y_2 completes the first cycle. Positive and negative predictive values can be computed after each cycle from Y_1/A and $(b-Y_2)/B$, respectively. The random samples generated by repeating the above cycle the desired number of times are then used to reconstruct the marginal posterior densities of each parameter and to find credible sets, marginal posterior means or medians, or other inferences.

4. Two diagnostic tests (conditional independence model)

The methods of the previous section can be extended to the situation where results of two diagnostic tests for the same disease are available on a randomly selected sample of subjects, where neither test can be considered a gold standard. There are unknown five parameters about which inference must be made: the population prevalence of CA dysfunction (π), and the sensitivity (S_1) and specificity (C_1) of the test1, and sensitivity (S_2) and specificity (C_2) of the Test2. Let U_1 be the observed number of positive test1 and test2 results, and U_2 be the observed number of positive test1 and negative test2 results, and U_3 the observed number of negative test1 and positive test2 results, and U_4 be the observed number of negative test1 and test2 results, in the sample of $U_1+U_2+U_3+U_4 = N$ subjects (Table a2).

Table a2: Observed data from two diagnostic tests, In the absence of a gold standard

		Test2		Total
		Positive	Negative	
Test1	Positive	U_1	U_2	U_1+U_2
	Negative	U_3	U_4	U_3+U_4
	Total	U_1+U_3	U_2+U_4	N

Let the unobserved latent data Y_1, Y_2, Y_3, and Y_4 represent the number of true positive subjects out of the observed cell values U_1, U_2, U_3 and U_4, respectively. Since any subject, whether truly possessing the disease in question or not, can test positively or negatively on each test, there are eight possible combinations.

Table a3: Likelihood contributions of all possible combinations of observed and latent data for the case of two independence diagnostic tests

No. of sub	Truth	Test1 result	Test2 result	Likelihood Contribution
Y_1	Positive	Positive	Positive	$\pi S_1 S_2$
Y_2	Positive	Positive	Negative	$\pi S_1(1-S_2)$
Y_3	Positive	Negative	Positive	$\pi(1-S_1)S_2$
Y_4	Positive	Negative	Negative	$\pi(1-S_1)(1-S_2)$
U_1-Y_1	Negative	Positive	Positive	$(1-\pi)(1-C_1)(1-C_2)$
U_2-Y_2	Negative	Positive	Negative	$(1-\pi)(1-C_1)C_2$
U_3-Y_3	Negative	Negative	Positive	$(1-\pi)C_1(1-C_2)$
U_4-Y_4	Negative	Negative	Negative	$(1-\pi)C_1C_2$

Note: The likelihood is proportional to the product of each entry In the last column of the table raised to the power of the corresponding entry In the first column of the table.

The likelihood function of the observed and latent data is given by (Table a3):

$$L(U_1,U_2,U_3,U_4,Y_1,Y_2,Y_3,Y_4|\pi,S_1,C_1,S_2,C_2)$$
$$=[\pi S_1S_2]^{Y_1}[\pi S_1(1-S_2)]^{Y_2}[\pi(1-S_1)S_2]^{Y_3}[\pi(1-S_1)(1-S_2)]^{Y_4}$$
$$[(1-\pi)(1-C_1)(1-C_2)]^{U_1-Y_1}[(1-\pi)(1-C_1)C_2]^{U_2-Y_2}[(1-\pi)C_1(1-C_2)]^{U_3-Y_3}[(1-\pi)C_1C_2]^{U_4-Y_4}$$

We used standard distributional families to represent our prior information. The choice of distributions discussed below is not unique and they may be replaced by other suitable densities, as needed. The prevalence is assumed to follow a beta prior distribution with parameters α and β, $\pi \sim \text{beta}(\alpha_\pi, \beta_\pi)$. The sensitivities and specificities are also assumed to have beta prior densities such that $S_j \sim \text{beta}(\alpha_{Sj}, \beta_{Sj})$, and $C_j \sim \text{beta}(\alpha_{Cj}, \beta_{Cj})$, j=1,2. The Gibbs sampler can again be used to construct the marginal posterior densities of all parameters of interest. For two independence diagnostic tests, the full conditional distributions are as follows:

$$Y_1|U_1,\pi,S_1,C_1,S_2,C_2 \sim Binomial(U_1, \frac{\pi S_1S_2}{\pi S_1S_2+(1-\pi)(1-C_1)(1-C_2)}) \quad \text{(app2.1)}$$

$$Y_2|U_2,\pi,S_1,C_1,S_2,C_2 \sim Binomial(U_2, \frac{\pi S_1(1-S_2)}{\pi S_1(1-S_2)+(1-\pi)(1-C_1)C_2}) \quad \text{(app2.2)}$$

$$Y_3|U_3,\pi,S_1,C_1,S_2,C_2 \sim Binomial(U_3, \frac{\pi(1-S_1)S_2}{\pi(1-S_1)S_2+(1-\pi)C_1(1-C_2)}) \quad \text{(app2.3)}$$

$$Y_4 | U_4, \pi, S_1, C_1, S_2, C_2 \sim Binomial(U_4, \frac{\pi\ (1-S_1)\ (1-S_2)}{\pi\ (1-S_1)\ (1-S_2)\ +(1-\pi) C_1 C_2}) \quad \text{(app2.4)}$$

$$\pi\ | U_1, U_2, U_3, U_4, Y_1, Y_2, Y_3, Y_4, \alpha_\pi, \beta_\pi, \sim Beta(Y_1+Y_2+Y_3+Y_4+\alpha_\pi, N\ -(Y_1+Y_2+Y_3+Y_4)+\beta_\pi) \quad \text{(app2.5)}$$

$$S_1 | Y_1, Y_2, Y_3, Y_4, \alpha_{S1}, \beta_{S1} \sim Beta(Y_1+Y_2+\alpha_{S1}, Y_3+Y_4+\beta_{S1}) \quad \text{(app2.6)}$$

$$S_2 | Y_1, Y_2, Y_3, Y_4, \alpha_{S2}, \beta_{S2} \sim Beta(Y_1+Y_3+\alpha_{S2}, Y_2+Y_4+\beta_{S2}) \quad \text{(app2.7)}$$

$$C_1 | U_1, U_2, U_3, U_4, Y_1, Y_2, Y_3, Y_4, \alpha_{C1}, \beta_{C1} \sim Beta(U_3+U_4-(Y_3-Y_4)+\alpha_{C1}, U_1+U_2-(Y_1+Y_2)+\beta_{C1}) \quad \text{(app2.8)}$$

$$C_2 | U_1, U_2, U_3, U_4, Y_1, Y_2, Y_3, Y_4, \alpha_{C2}, \beta_{C2} \sim Beta(U_2+U_4-(Y_2-Y_4)+\alpha_{C2}, U_1+U_3-(Y_1+Y_3)+\beta_{C2}) \quad \text{(app2.9)}$$

5. Two diagnostic tests (conditional dependence model)

Assume that we have results from two different dichotomous tests T_j, j=1,2, from a sample of N subjects such that a positive result on the jth test is denoted by T_j=1 and a negative result by T_j=0. Let D denote the latent true disease status such that D=1 among diseased subjects and D=0 among nondiseased subjects. To model the conditional dependence between two diagnostic tests recommended by Dendukuri et al., the conditional dependence between tests may be estimated using a measure such as the covariance between tests within each disease class. We denote the covariance between the two tests among the diseased and nodiseased subjects as cov_s and cov_c, respectively. Here, cov_s=P(T_1=1, T_2=1|D=1)-S_1S_2, and cov_c= P(T_1=0, T_2=0|D=0)-C_1C_2.

Table a4: Likelihood contributions of all possible combinations of observed and latent data for the case of two dependence diagnostic tests

No. of sub	Truth	Test1 result	Test2 result	Likelihood Contribution
Y_1	Positive	Positive	Positive	$\pi(S_1S_2+cov_s)$
Y_2	Positive	Positive	Negative	$\pi(S_1(1-S_2)-cov_s)$
Y_3	Positive	Negative	Positive	$\pi((1-S_1)S_2-cov_s)$
Y_4	Positive	Negative	Negative	$\pi((1-S_1)(1-S_2)+cov_s)$
U_1-Y_1	Negative	Positive	Positive	$(1-\pi)((1-C_1)(1-C_2)+cov_c)$
U_2-Y_2	Negative	Positive	Negative	$(1-\pi)((1-C_1)C_2-cov_c)$
U_3-Y_3	Negative	Negative	Positive	$(1-\pi)(C_1(1-C_2)-cov_c)$
U_4-Y_4	Negative	Negative	Negative	$(1-\pi)(C_1C_2+cov_c)$

Note: The likelihood is proportional to the product of each entry In the last column of the table raised to the power of the corresponding entry In the first column of the table.

The likelihood function of the observed and latent data is given by (Table a4):

$$L(U_1,U_2,U_3,U_4,Y_1,Y_2,Y_3,Y_4 \mid \pi, S_1, C_1, S_2, C_2, cov_s, cov_c)$$
$$= [\pi(S_1S_2+cov_s)]^{Y_1}[\pi(S_1(1-S_2)-cov_s)]^{Y_2}[\pi((1-S_1)S_2-cov_s)]^{Y_3}[\pi((1-S_1)(1-S_2)+cov_s)]^{Y_4}$$
$$[(1-\pi)((1-C_1)(1-C_2)+cov_c)]^{U_1-Y_1}[(1-\pi)((1-C_1)C_2-cov_c)]^{U_2-Y_2}[(1-\pi)(C_1(1-C_2)-cov_c)]^{U_3-Y_3}[(1-\pi)(C_1C_2+cov_c)]^{U_4-Y_4}$$

We used standard distributional families to represent our prior information. The choice of distributions discussed below is not unique and they may be replaced by other suitable densities, as needed. The prevalence is assumed to follow a beta prior distribution with

parameters α and β, π~beta(α_π, β_π). The sensitivities and specificities are also assumed to have beta prior densities such that S_j~beta(α_{Sj}, β_{Sj}), and C_j~beta(α_{Cj}, β_{Cj}), j=1,2. The feasible range of the covariance is determined by the sensitivities among the disease subjects and the specificities among the nondiseased subjects. The covariance parameters are taken to have uniform prior distribution, covs~uniform(0,min(S_1,S_2)-S_1S_2) and covc~uniform(0, min(C_1,C_2) -C_1C_2), where min(a,b) is the minimum of a and b. The Gibbs sampler can again be used to construct the marginal posterior densities of all parameters of interest. For two independence diagnostic tests, the full conditional distributions are as follows:

$$Y_1|U_1,\pi,S_1,C_1,S_1,C_1,\mathrm{cov}_s,\mathrm{cov}_c \sim Binomial(U_1,\frac{\pi\ (S_1S_2+\mathrm{cov}_s)}{\pi\ (S_1S_2+\mathrm{cov}_s)+(1-\pi)((1-C_1)(1-C_2)+\mathrm{cov}_c)}) \quad \text{(app2.10)}$$

$$Y_2|U_2,\pi,S_1,C_1,S_1,C_1,\mathrm{cov}_s,\mathrm{cov}_c \sim Binomial(U_2,\frac{\pi\ (S_1(1-S_2)\ -\mathrm{cov}_s)}{\pi\ (S_1(1-S_2)\ -\mathrm{cov}_s)+(1-\pi)((1-C_1)C_2-\mathrm{cov}_c)}) \quad \text{(app2.11)}$$

$$Y_3|U_3,\pi,S_1,C_1,S_1,C_1,\mathrm{cov}_s,\mathrm{cov}_c \sim Binomial(U_3,\frac{\pi\ ((1-S_1)\ S_2-\mathrm{cov}_s)}{\pi\ ((1-S_1)\ S_2-\mathrm{cov}_s)+(1-\pi)(C_1(1-C_2)-\mathrm{cov}_c)}) \quad \text{(app2.12)}$$

$$Y_4|U_4,\pi,S_1,C_1,S_1,C_1,\mathrm{cov}_s,\mathrm{cov}_c \sim Binomial(U_4,\frac{\pi\ ((1-S_1)\ (1-S_2)\ +\mathrm{cov}_s)}{\pi\ ((1-S_1)\ (1-S_2)\ +\mathrm{cov}_s)+(1-\pi)(C_1C_2+\mathrm{cov}_c)}) \quad \text{(app2.13)}$$

$$\pi\ |U_1,U_2,U_3,U_4,Y_1,Y_2,Y_3,Y_4,\alpha_\pi,\beta_\pi, \sim Beta(Y_1+Y_2+Y_3+Y_4+\alpha_\pi, N\ -(Y_1+Y_2+Y_3+Y_4)+\beta_\pi) \quad \text{(app2.14)}$$

$$S_1|Y_1,Y_2,Y_3,Y_4,\alpha_{S1},\beta_{S1} \sim Beta(Y_1+Y_2+\alpha_{S1},Y_3+Y_4+\beta_{S1}) \quad \text{(app2.15)}$$

$$S_2|Y_1,Y_2,Y_3,Y_4,\alpha_{S2},\beta_{S2} \sim Beta(Y_1+Y_3+\alpha_{S2},Y_2+Y_4+\beta_{S2}) \quad \text{(app2.16)}$$

$$C_1|U_1,U_2,U_3,U_4,Y_1,Y_2,Y_3,Y_4,\alpha_{C1},\beta_{C1} \sim Beta(U_3+U_4-(Y_3-Y_4)+\alpha_{C1},U_1+U_2-(Y_1+Y_2)+\beta_{C1}) \quad \text{(app2.17)}$$

$$C_2|U_1,U_2,U_3,U_4,Y_1,Y_2,Y_3,Y_4,\alpha_{C2},\beta_{C2} \sim Beta(U_2+U_4-(Y_2-Y_4)+\alpha_{C2},U_1+U_3-(Y_1+Y_3)+\beta_{C2}) \quad \text{(app2.18)}$$

$$\mathrm{cov}_s \sim uniform(0,\min(S_1,S_2)-S_1S_2) \quad \text{(app2.19)}$$

$$\mathrm{cov}_c \sim uniform(0,\min(C_1,C_2)-C_1C_2) \quad \text{(app2.20)}$$

Gibbs sampling is used to sample in turn from distribution app2.10 to distribution app2.20 in a similar fashion to the procedure used for the case of one diagnostic test outlined previously. The positive and negative predictive values for each cycle of the Gibbs algorithm

are again obtained directly from the relevant fractions of the true positive or negative subjects in each cell of the 2 by 2 table to the total observed number of subjects in that cell. Throughout, the Gibbs sampler was run for 100,000 cycles, the first 10,000 to assess convergence and the last 90,000 for inference. Each analysis was repeated from several different starting values, and convergence was assumed only if all runs provided very similar posterior distributions. Convergence of the algorithm here appeared to occur within the first 100-200 cycles, as evidenced by the monitoring of selected percentiles of the posterior samples. In general, the rate of convergence will depend on the starting values and the particulars of the data set and prior distributions. A computer program written in S-PLUS implementing all of the methods described in this paper is available from the first author (albert.tang@163.com).

6. WinBUGS program code for two independence diagnostic test

```
//
model;
{     pi ~ dbeta(api,bpi)
      sen1 ~ dbeta(as1,bs1)
      spe1 ~ dbeta(ac1,bc1)
      sen2 ~ dbeta(as2,bs2)
      spe2 ~ dbeta(ac2,bc2)

      api<-x[1]document+x[10]+x[11]+x[4]+api0
      bpi<-n-(x[1]+x[10]+x[11]+x[4])+bpi0

      as1<-x[1]+x[10]+as10
      bs1<-x[11]+x[4]+bs10
      ac1<-u[11]+u[4]-(x[11]+x[4])+ac10
      bc1<-u[1]+u[10]-(x[1]+x[10])+bc10

      as2<-x[1]+x[11]+ as20
      bs2<-x[10]+x[4]+ bs20
      ac2<-u[10]+u[4]-(x[10]+x[4])+ac20
      bc2<-u[1]+u[11]-(x[1]+x[11])+bc20

      p[1] <- pi * sen1 * sen2/(pi * sen1 * sen2 + (1 - pi) * (1 - spe1) * (1 - spe2))
      p[10] <- pi * sen1 * (1-sen2)/(pi * sen1 * (1-sen2) + (1 - pi) * (1 - spe1) * spe2)
      p[11] <- pi * (1-sen1) *sen2/(pi * (1-sen1) *sen2 + (1 - pi) *    spe1 * (1-spe2))
      p[4] <- pi * (1-sen1) * (1-sen2)/(pi * (1-sen1) * (1-sen2) + (1 - pi) * spe1 * spe2)

     for(i in 1:N)
      { x[i] ~ dbin(p[i],u[i])
      }
}
List(data format...)
//
```

Reference:

1. Branscum AJ, Gardner IA, Johnson WO (2005) Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. Preventive veterinary medicine 68: 145-163.
2. Spallone V, Ziegler D, Freeman R, Bernardi L, Frontoni S, et al. (2011) Cardiovascular autonomic neuropathy in diabetes: clinical impact, assessment, diagnosis, and management. Diabetes Metab Res Rev.
3. Pop-Busui R (2010) Cardiac autonomic neuropathy in diabetes: a clinical perspective. Diabetes care 33: 434-441.
4. Vinik AI, Ziegler D (2007) Diabetic cardiovascular autonomic neuropathy. Circulation 115: 387-397.
5. Straub RH, Lang B, Palitzsch KD, Scholmerich J (1997) Estimation of the cut-off value in cardiovascular autonomic nervous function tests: not-age-related criteria or the age-related 5th percentile. Journal of diabetes and its complications 11: 145-150.
6. Neil HA, Thompson AV, John S, McCarthy ST, Mann JI (1989) Diabetic autonomic neuropathy: the prevalence of impaired heart rate variability in a geographically defined population. Diabetic medicine : a journal of the British Diabetic Association 6: 20-24.
7. Risk M, Bril V, Broadbridge C, Cohen A (2001) Heart rate variability measurement in diabetic neuropathy: review of methods. Diabetes technology & therapeutics 3: 63-76.
8. Ziegler D, Dannehl K, Muhlen H, Spuler M, Gries FA (1992) Prevalence of cardiovascular autonomic dysfunction assessed by spectral analysis, vector analysis, and standard tests of heart rate variation and blood pressure responses at various stages of diabetic neuropathy. Diabetic medicine : a journal of the British Diabetic Association 9: 806-814.
9. Dendukuri N, Joseph L (2001) Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics 57: 158-167.
10. Care A, Catalucci D, Felicetti F, Bonci D, Addario A, et al. (2007) MicroRNA-133 controls cardiac hypertrophy. Nat Med 13: 613-618.
11. Thayer JF, Merritt MM, Sollers JJ, 3rd, Zonderman AB, Evans MK, et al. (2003) Effect of angiotensin-converting enzyme insertion/deletion polymorphism DD genotype on high-frequency heart rate variability in African Americans. The American journal of cardiology 92: 1487-1490.