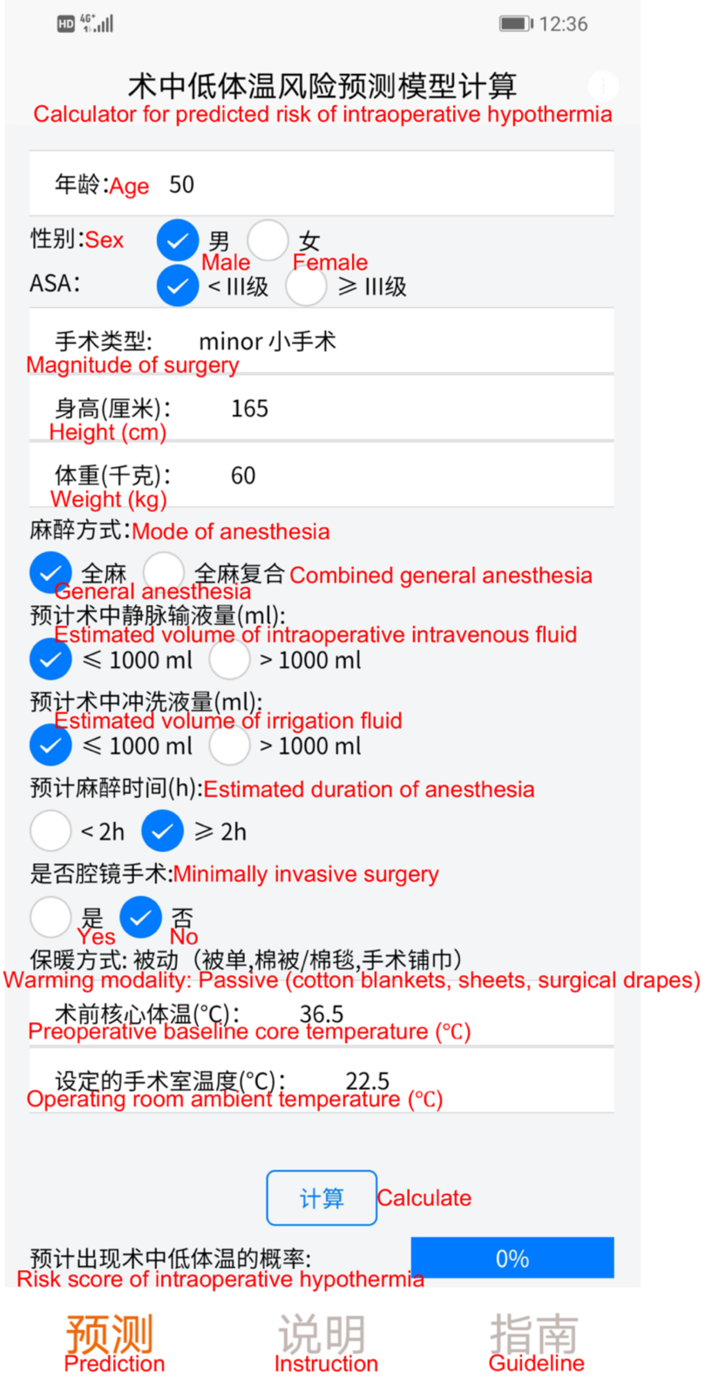
Supplemental Digital Content



术中低体温风险预测模型计算
Calculator for predicted risk of intraoperative hypothermia

年龄:Age 50

性别:Sex ☑ 男 Male ○ 女 Female

ASA: ☑ < III级 ○ ≥ III级

手术类型: minor 小手术
Magnitude of surgery

身高(厘米): 165
Height (cm)

体重(千克): 60
Weight (kg)

麻醉方式:Mode of anesthesia

☑ 全麻 General anesthesia ○ 全麻复合 Combined general anesthesia

预计术中静脉输液量(ml):
Estimated volume of intraoperative intravenous fluid

☑ ≤ 1000 ml ○ > 1000 ml

预计术中冲洗液量(ml):
Estimated volume of irrigation fluid

☑ ≤ 1000 ml ○ > 1000 ml

预计麻醉时间(h):Estimated duration of anesthesia

○ < 2h ☑ ≥ 2h

是否腔镜手术:Minimally invasive surgery

○ 是 Yes ☑ 否 No

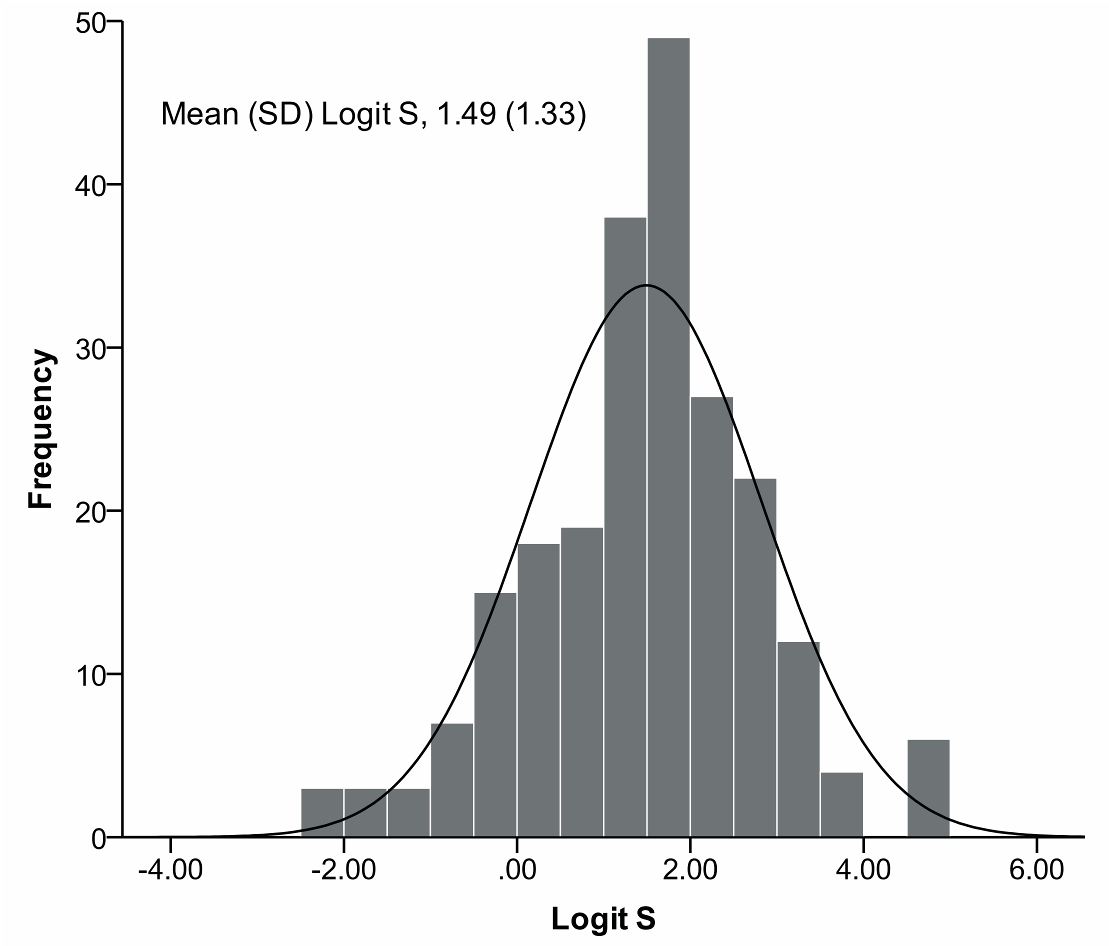
保暖方式: 被动（被单,棉被/棉毯,手术铺巾）
Warming modality: Passive (cotton blankets, sheets, surgical drapes)

术前核心体温(°C): 36.5
Preoperative baseline core temperature (℃)

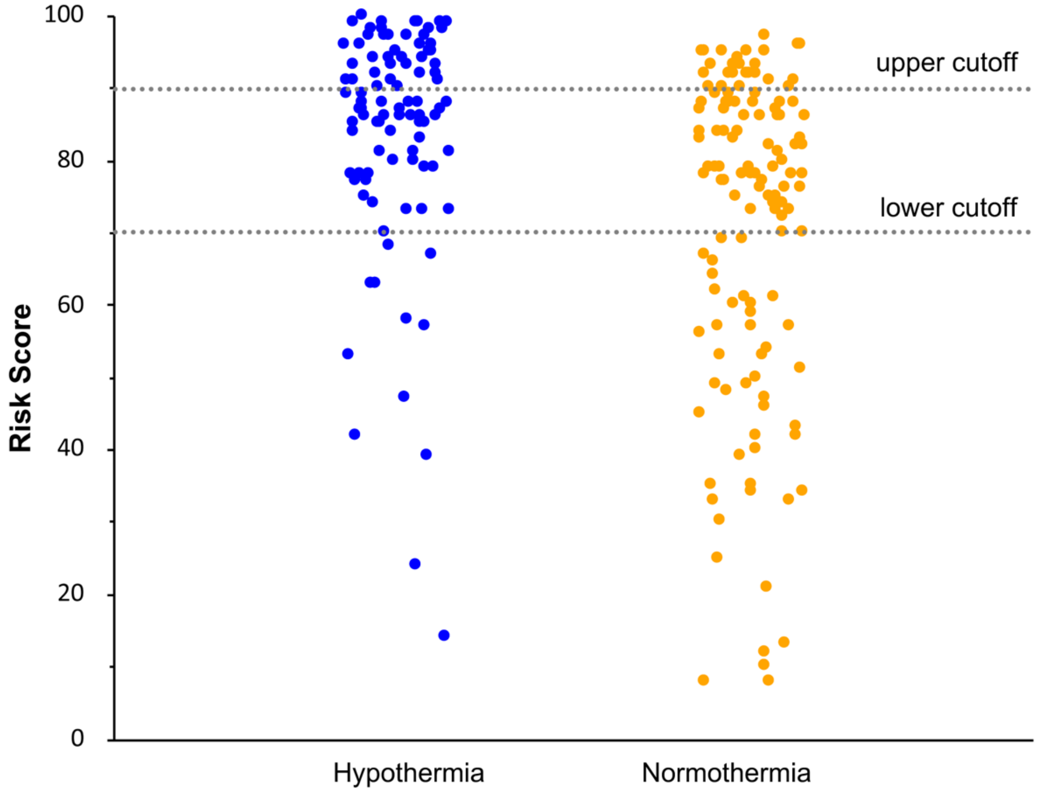
设定的手术室温度(°C): 22.5
Operating room ambient temperature (℃)

预计出现术中低体温的概率: 0%
Risk score of intraoperative hypothermia

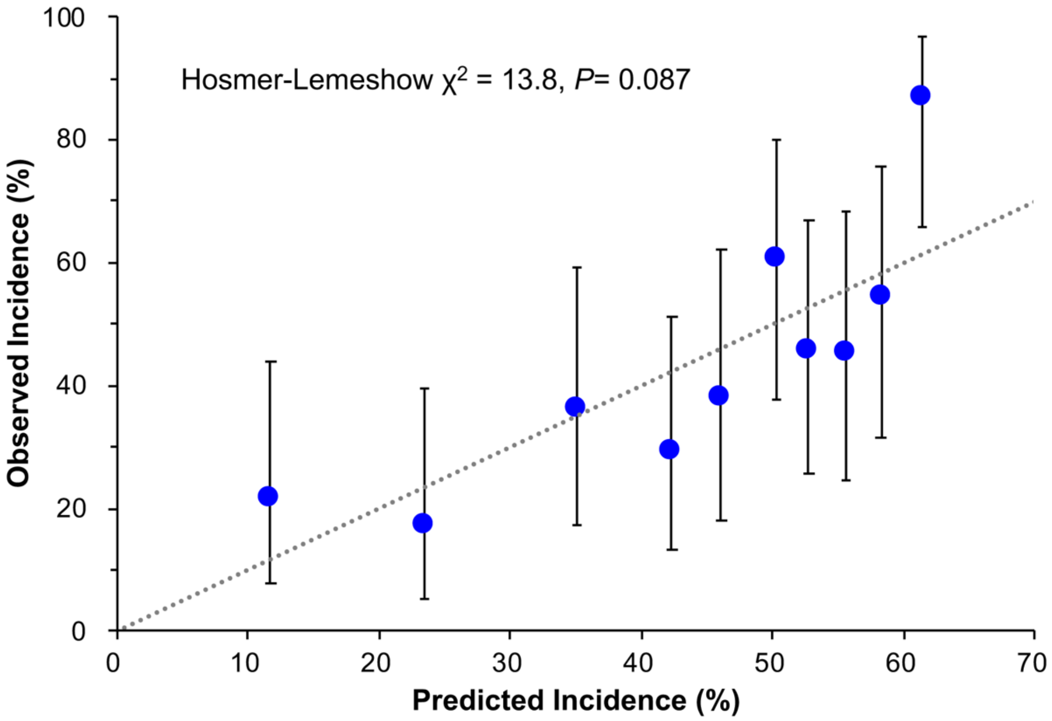
eFigure 1. Intraoperative Hypothermia Predictor APP based on the risk prediction model. A screenshot showing the mobile application interface of risk score calculation. It also contains two other interfaces providing instructions and guidelines. The red words are English translations of each item. APP indicates application; ASA, American Society of Anesthesiologists.



eFigure 2. Frequency histogram for Logit S among 227 participants with a mean ± SD Logit S of 1.49 ± 1.33. Logit S = $\log\frac{score\%}{1-score\%}$. The normal distribution curve represents the fitting curve for Logit S distribution. SD indicates standard deviation.



eFigure 3. Distribution of the risk scores by the results of perioperative temperature monitoring in 227 participants. Gray dashed lines represent lower and upper cutoffs set at 70 and 90, respectively. Jitter was introduced to avoid overlapping.



eFigure 4. Calibration plot for the intraoperative hypothermia prediction model showing the predicted versus the observed incidence for 10 groups in the whole cohort (N=227). A Hosmer-Lemeshow χ^2 of 13.8 suggests good calibration, with *P* value of 0.087. The error bars represent 95% confidence intervals. The gray dashed line represents an excellent model whose expected incidence is the same as the observed incidence.

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	2
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	3-4
	4	Study objectives and hypotheses	4
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	4
Participants	**6**	Eligibility criteria	4-5
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	4
	8	Where and when potentially eligible participants were identified (setting, location and dates)	4
	9	Whether participants formed a consecutive, random or convenience series	4
Test methods	**10a**	Index test, in sufficient detail to allow replication	6-7
	10b	Reference standard, in sufficient detail to allow replication	6
	11	Rationale for choosing the reference standard (if alternatives exist)	6
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	8
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	7
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	6
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	6
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	8-9
	15	How indeterminate index test or reference standard results were handled	7
	16	How missing data on the index test and reference standard were handled	7
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	8-9, 15
	18	Intended sample size and how it was determined	9
RESULTS			
Participants	**19**	Flow of participants, using a diagram	9, 10
	20	Baseline demographic and clinical characteristics of participants	10, 22-24
	21a	Distribution of severity of disease in those with the target condition	10
	21b	Distribution of alternative diagnoses in those without the target condition	10
	22	Time interval and any clinical interventions between index test and reference standard	6
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	10-11
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	11
	25	Any adverse events from performing the index test or the reference standard	6
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	15
	27	Implications for practice, including the intended use and clinical role of the index test	16
OTHER INFORMATION			
	28	Registration number and name of registry	Not applicable.
	29	Where the full study protocol can be accessed	Not applicable.
	30	Sources of funding and other support; role of funders	16

STARD 2015

AIM

STARD stands for “Standards for Reporting Diagnostic accuracy studies”. This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or “2x2” table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.