CUMCM-2025 Problem C

Time Determination of NIPT and

Abnormality Discrimination of the Fetus

Non-invasive Prenatal Test (NIPT) is a prenatal test technique that collects maternal peripheral blood, detects fetal cell-free DNA fragments, and analyzes fetal chromosomes for abnormalities. Its purpose is to determine the health status of the fetus through early detection. According to clinical experience, fetuses with deformities mainly suffer from Down's syndrome, Edward's syndrome, and Patau's syndrome, which can be detected by using the cell-free DNA fragments of chromosome 13,18 and 21, respectively. The accuracy of NIPT is practically judged by the fetal fraction ("FF of Chr." for short) in sex chromosome (Chr.Y for male fetus and Chr.X for female fetus) from the 10th week to the 25th week of pregnancy. The NIPT results can be considered as accurate if the FF of Chr.Y for male fetus is higher than level 4% and the FF of Chr. X for female fetus is normal. It could be difficult to ensure the accuracy of the results when the FF of Chr.Y for male fetus is less than 4%. In practice, NIPT should be implemented as early as possible to reduce the risk of delayed treatment when abnormality is detected. Early detection within 12 weeks of pregnancy has a lower risk, mid-term detection between 13 and 27 weeks of pregnancy indicates a high risk, and the risk of late detection after 28 weeks will be extremely high.

Previous studies have shown that the FF of Chr.Y of male fetus is highly related to the weeks of pregnancy and the body mass index (BMI) of pregnant women. The time points of NIPT are usually determined by grouping the BMI value of pregnant women (for example: [20,28, [28,32), [32,36), [36,40) and above 40). Due to the individual heterogeneity in age, BMI, pregnancy status and other factors among pregnant women, the accuracy of NIPT with simple empirical grouping and unified testing time points for all pregnant women is greatly affected in practice. Therefore, in order to reduce the potential risks of delayed treatment for some pregnant women with unhealthy fetus, it is necessary to group pregnant women reasonably according to BMI and determine the best time node of NIPT for different groups.

To investigate the appropriate testing time of NIPT for various groups of pregnant women and analyze the accuracy of the test, the NIPT dataset of pregnant women (mostly with high BMI) is given in data.xlsx. In the practical tests, it is often cases that the test results fail (for example: too early test time and uncertain factors, etc). Furthermore, in order to increase the reliability of the test results, some pregnant women have multiple blood draws and tests or one blood draw and multiple tests. Please establish mathematical models to study the following problems with the dataset in data.xlsx

Problem 1 Make correlation analysis between fetal FF of Chr. Y and give some indices such as gestational weeks and BMI of pregnant women, and provide corresponding relationship models and test their significance.

Problem 2 Clinical evidence shows that the BMI of pregnant women with male fetus is the main factor affecting the earliest time of the fetal FF of Chr.Y reaches or exceeds level 4%. Please group reasonably the BMI of pregnant women with male fetus, give out the BMI interval and optimal NIPT time determination for each group, so as to minimize the potential risk of pregnant women, and analyze the impact of the test errors on the results.

Problem 3 The earliest time of the fetal FF of Chr.Y reaches or exceeds level 4% is influenced by

various factors such as height, weight, age, etc. Please consider these factors, test errors, and the proportion of fetal FF of Chr.Y reaching or exceeding level 4%, give a reasonable grouping and the optimal NIPT time determination for each group based on the BMI of pregnant women with male fetus, so as to minimize the potential risks of pregnant women, and analyze the impact of the test errors on the results.

Problem 4 Since neither pregnant women nor female fetuses carry chromosome Y, it is important to determine whether the female fetus is abnormal. Based on the results of aneuploidy of chromosomes 21, 18, and 13 (AB column) in pregnant women, taking into account factors such as the FF of Chr. X, the Z-score, GC content of the aforementioned chromosomes, number of reads and related ratios, and BMI etc., please give out a discriminant method of the abnormality of female fetuses.

Appendix 1 Description of column data in data.xlsx

Col	Description	Col	Description
A	Sample number	Q	Z-score of chromosome 13
B	Code of pregnant women	R	Z-score of chromosome 18
C	Age of pregnant women	S	Z-score of chromosome 21
D	Height of pregnant women	T	Z-score of chromosome X
E	Weight of pregnant women	U	Z-score of chromosome X（it is blank for test data-female fetus in data.xlsx）
F	Last menstrual period	V	FF of Chr. Y：the proportion of the cell- free DNA fragments of chromosome Y（it is blank for test data-female fetus in data.xlsx）
G	IVF pregnancy	W	FF of Chr.X：its value is estimated through bioinformatics and data analysis under certain assumptions, and negative values may occur
H	Test time	X	GC content of chromosome 13
I	Number of blood draws	Y	GC content of chromosome 18
J	Pregnant women's gestational age at the time of this test（weeks+days）	Z	GC content of chromosome 21
K	BMI of pregnant women	AA	Filter rate
L	Raw reads number	AB	More aneuploidy（Chrs.13，18，21），Blank means no abnormalities
M	Align rate	AC	Number of pregnancies
N	Duplication rate	AD	Gravida para
O	UReads number	AE	Fetal health（outcome after birth of baby）
P	GC content：the proportion of nucleotides G (guanine) and C (cytosine) in a sequence is an important index for evaluating the quality of sequencing data. The normal GC content range is 40%-60%, and high, low, or abnormally distributed GC content may indicate some problems with sequencing quality		Blank or NA: Not tested (except for columns AB and U, V for test data-female fetus in data.xlsx)

Appendix 2 Z-score

Z-score can be expressed by

$$Z = \frac{X - \mu}{\sigma}$$

where X is the relative counting ratio of the target chromosome in the sample to be tested, μ and σ are the mean and the standard deviation respectively of the chromosome counting ratio in the normal control population. In NIPT, Z-score analysis is usually used for statistical discrimination of common chromosomal aneuploidy detection. Chromosomal aneuploidy is usually defined as the presence of one or three copies of the chromosome, with two copies being normal, and the number of reads collected for each chromosome is proportional to its length.