

National Focused Open Data (Health Checkup Information) User Manual (ver 4.0)



National Health Insurance

Corporation Big Data Operation Office

목 차

제#io절 개요

1. What is ÿNational Focused Open Dataÿ? One

go, concept 1

me. Open Purpose 1

2. Progress of the development of ÿNational Focused Open Dataÿ 1

제 절 국가중점 개방데이터 건강검진정보 국생 방향성

1. National-focused open data selection criteria 2

go. Safe data-oriented openness (Safety) - Focus on personal information protection $\ensuremath{\mathbf{2}}$

1) Excluding personally identifiable information that can identify an individual by itself 2

2) Application of personal information de-identification processing technique 2

me. Utilization-oriented data openness - user-tailored data openness 2 $\,$

All. Objective and Reliable Data Reliability 3

2. Open Dataset Generation Type 3

go. Open Dataset Creation Method and Scope 3

me. Open dataset update cycle 3

All. Open data set provision type 3

제 교절 국가중점 개방데이터 건강검진정보 상세

ÿHealth checkup informationÿ 4

1. Concept and composition 4

1) Concept 4

2) Configuration 4

2. Description of open items 4

3. Precautions when using health checkup information 8 $\,$

Machine ¹	Translated by Google
Г	
	게 저 그가주저 게바레이다 경가경지저나 다오르드 바버
	제대절 국가중점 개방데이터 건강검진정보 다운로드 방법
	1. Introduction of how to use the National Health Insurance Data Sharing Service (NHISS) webpage 9
	폐···절 사용자 지원 사항
	1. Provide national-focused open data utilization guides and examples 11
	2. Support for online inquiry function using the website for opening national-focused open data 11
	3. Support for providing demo data before downloading country-focused open data 11
	Q&A를Q8답하기



What is ÿNational Focused Open Dataÿ?

go. concept

Nationally-focused open data is open to the public focusing on items that require a high demand for openness from the private sector among the data held by the National Health Insurance Corporation in line with the government 3.0 policy. The nationally-focused open data is the medical history information, drug prescription information, and health checkup information of the national health insurance subscribers accumulated as the NHIS plays the role of the national health insurance subscriber. In order to safely open data, the Corporation excluded or masked personal information and sensitive data, and provided advice regarding the safety of open data through an external expert. The target year of the provided data is from 2002 to 2015, and it is planned to continuously expand the target period in the future. It is hoped that it will contribute to job creation for the public by revitalizing related industries through the opening of national-focused open data.

me. open purpose

첫째 seeks to guarantee the public's right to use public data based on the law and to satisfy the public's right to know about major health and medical information collected, held, and managed by the National Health Insurance Corporation. we war property we want to read the local property in the health information-related industry using open national by opening individual-centered treadistributed tails and drug prescription information to the general public, it is intended to induce medical institutions to provide high-quality medical services. Additionally, the purpose is to strengthen national competitiveness in the health and medical industry due to the development of health information-related industries and to create jobs through new business development.

Progress of development of ÿNational Focused Open Dataÿ



Starting with the establishment of an initial national health information DB by collecting health insurance information for 50 million people across the country for 9 years (2002-2010) in 2011, MOU with 16 institutions including Seoul National University Big Data Center was signed in 2013. A pilot research project was promoted. In 2014, in order to promote data disclosure, consulting on the development of an appropriate fee system for public health information DB disclosure, the ÿNational Health Information Data Provision Operational Regulation (Regulation No. 130)ÿ was enacted. In January 2015, application for use of sample cohort DB and public health A dedicated website was developed to support information DB researchers. In December 2015, for the purpose of opening to the private sector, the government-focused open data consisting of medical history information, drug prescription information, and health examination information data sets were established and provided, and additional data 2014-2015 data were updated in December 2016 (current), was continuously provided.

🤊 국민건강정보

Analyze the data held by the National Health Insurance Corporation as big data and open the data.

Data discovery and creation an adversarially built population

DB includes information such as qualifications and insurance premiums, health checkup results and questionnaires, detailed treatment history (hospital/clinic usage history), and long-term care for the elderly. The total number of tables is 3,544 and the capacity is about 21,236 GB.

Section 2 Direction of National Focused Open Data Composition

Criteria for selection of country-focused open data

go. Safe data-oriented openness (Safety) - focus on personal information protection $% \left(1\right) =\left(1\right) \left(1$

1) Excluding personally identifiable information that can identify an individual by itself

Personal identification information that can identify an individual (ex. unique identification information (resident number, national health insurance qualification number, etc.), information that can easily identify an individual (name, phone number, address, date of birth, photo, etc.), etc. was excluded from the country-focused open data items.

2) Application of personal information de-identification processing technique

Data with the possibility of re-identification by pre-filtering whether or not personally identifiable information is included in combination with other information ldentification was excluded either by exclusion from the subject of disclosure or by applying a de-identification processing technique suitable for individual items.

구분	내용
표본의 추출	Randomly selected 1 million patients for treatment and health checkup by year ÿ 1 million patients for treatment, prescription medication, and health checkup were randomly selected for each year
자료별 조합제한	This data is a cross-sectional data and provided with different personal serial numbers and claim serial numbers for each DB (not linked)
개인식별자 제거 · Reside	nt registration number (13) ÿ Personal serial number (8 digits)
범주화	Age grouping ÿ Age ÿ Age group (units of 5 years old) ÿ 85 years old and older are categorized as '85 years old or older'
데이터 마스킹	• Large classification of sensitive disease D, O, P, X, Y codes (5 types, 114) (Example: D***) ÿ D (7 codes): benign neoplasms of male reproductive organs ÿ O (81 codes): pregnancy, childbirth and postpartum, ectopic pregnancy, etc. ÿ P (4 codes): Other congenital infections and parasitic diseases ÿ X (1 code): accidental poisoning and exposure to alcohol ÿ Y (2 codes): special purpose code, sexual abuse by force, etc.
최상위 지역코드 제공	Provided limited to provincial and provincial codes in consideration of the recognition of samples living in small areas (17 provincial and provincial units provided)

me. Utilization-oriented data openness - user-tailored data opening

For national-focused open data, a demand survey related to open data items and methods of provision was conducted, and the open items and provision types were formed by reflecting the results of the demand survey. However, some of the items requested through the demand survey were classified as sensitive data and the opening was restricted.

All. Objective and reliable data openness (Reliability)

In order to ensure that nationally-focused open data can be trusted and used in various fields, items with low reliability or difficult to verify data values from the time of discovery of open data were excluded from open items. For example, items directly entered by the examinee among the health checkup information questionnaire items are incorrect in the data value.

Efforts were made to ensure objectivity and reliability through the process of refining the flow .

Open Dataset Generation Form

go. Open dataset creation method and scope

The national-focused open data is information on the medical treatment history, drug prescription history, and health check-up results of 1 million national health insurance subscribers from 2002 to 2015. A random screening technique was applied to screening 1 million people, and this manual is composed of information on health checkup information.

For health checkup information, one million national health insurance subscribers who underwent a health checkup in the given year are randomly selected, and basic information and checkup result information of the selected subscribers are extracted through the item selection process to construct a national health information dataset, did

me. Open dataset update cycle

All open data sets of country-focused open data are updated annually by adding data from two years ago in December every year. For example, data for 2014 is added in December 2016.

All. Open data set provision form

The data structure (format) of the dataset is CSV (Comma Separated Value), and it is divided by year for the convenience of data use and download.

Data is provided in a formatted format.

Section 3 State-focused open data details

ÿHealth checkup informationÿ

1. Concept and composition

1) Concept

Health check-up information refers to the number of employees enrolled in the National Health Insurance system, dependents over 40 years of age, and local insured persons who are the head of the household, from 2002 to 2015 (14 years).

The results of the ÿGeneral Medical Examinationÿ of local insured persons over the age of 40 and those who have reached the age of 40 and 66 among those subject to these general health examinations will receive

This is the result of the "Life Transition Period Health Checkup".

구분	검진대상	검진주기	
TE	차	차	100구기
General health checkup	Local head of household, member of household over 40 years of age Employed subscribers, workers over 40 years of age	1st examination result Suspected disease (high risk group)	Once in 2 years (Non-office positions every year)
life transition period medical examination	Subjects for general health examination who have reached the age of 40 Subjects for general health examination who have reached the age of 66	All 1st examinees	Applicable age

ÿ Workplace subscribers and local householders are eligible for health checkup without age restrictions. However, in this health examination information, the examination results of those under the age of 20 were excluded.

2) Configuration

With a total of 34 variables, the serial number of the subscriber and **种品和**加**尼**

basic information such as gender, age, residence, city and province code ÿ건강검진결과 및 문진정

🗷 Body size information such as body, weight, waist circumference, and pathological examination results such as blood pressure, blood sugar, cholesterol, urine protein, and gamma GP.

It consists of the results of other diagnostic tests such as oral examinations, as well as the results of questionnaires about drinking and smoking.

2. Description of open items

			제공항목	속성	정보	
NO	표준 항목명	영문명	설명	표현형식 단 위	예시	비고
1	base year do	HCHK_YEAR	Provide the base year of the information	YYYY	2009	
2	subscriber serial number like	IDV_ID	Serial number assigned to the subscriber 1 – 1,000,000	N	1	
3	sex nose de	SEX	provide the gender of the subject of the information; Gender: 1 (male), 2 (female)	N	1	•
4 ages	code (5 three units)	AGE_GROUP	• In the base year, the age of the examinees is grouped by age 5 (categorization) code separated by Grouped by 5-year-olds, 85+ grouped by 85+ From 2002 to 2013 그룹 연령대 그룹 연령대	11	·	

		1	1 1				1	11	1	ı «		
			1	20-24 years o		8	55-59 years old					
			2	25-29 years o		9	60-64 years old					
			3	30-34 years o	ild	10	65-69 years old					
			4	35-39 years o	ild	11	70-74 years old					
			5	40-44 years o	ıld	12	75-79 years old					
			6	45-49 years o	ıld	13	80-84 years old					
			7	50-54 years o	ild	14	85+					
			• Since 2014					2				
			그룹	연령대		그룹	연령대					
			1	0-4 years old		10	45-49 years old					
			2	5 to 9 years old		11	50-54 years old					
			3	10-14 years old		12	55-59 years old					
			4	15-19 years old		13	60-64 years old					
			5	20-24 years old		14	65-69 years old					
			6	25-29 years old		15	70-74 years old					
			7	30-34 years old		16	75-79 years old					
			8	35-39 years old		17	80-84 years old					
			9	40-44 years old		18	85+					
			As Sejo	the examinee's resident ong Special Self-Governing not exist in the data up to		r incorporated fro	m 2012, 2011					
					코드명	시도명	코드명		시도명			
					11 Seoul		42 Gang	wo 1-do				
				26 Busan		43 Chung	cheongbuk-do					
5	Sidoko			BEEN	27 Daegu Me	tropolitan City	44 Chung	gcheongnam-do		N	26	
	de	SEEN	28 Incheon I	letropolitan City	45 Jeolla	bu t-do			25			
			29 Gwangju		46 Jeolla	nam-do						
			30 Daejeon		47 Gyeor	ngsangbuk-do						
			31 Ulsan Me	ropolitan City	48 Gyeor	ngsangnam-do						
			36 Sejong S	ecial Self-Governing City 49 Jeju	Special Self-Gov	erring Province						
			41 Gyeonggi	do								
					1	1						
	kidney		• The examinee	's height (in units of 5Cl	M)							
6	(5cm step	HEIGHT		04CM -> 100CM				N/Cm	140			
	stomach)		y LA) 100~					1				
7	weight	WEIGHT	The examinee	's weight (in units of 5K	G)			100	45			
,	(5Kg only	WEIGHT	ÿ Example)	25~29KG -> 25KG				N/Kg	45			
	and the state of t											
	waist two		Examinee's warened to be a second to be a seco	aist circumference								
8	Re	WAIST	1	umference was added a				N/Cm	82			
			processed and s		,19							
			• Vision in the	vaminee's left ava				1				
9	Vision	SIGHT_LEFT		xaminee's left eye ressed as a value between	0.1 and 2.5. V	isual acuity below	v 0.1 is 0.1, and blindness is	N	0.5			
	(left)			It is expressed as a value between 0.1 and 2.5. Visual acuity below 0.1 is 0.1, and blindness is is written as 9.9								
3			1					1				
10	Vision	SIGHT_RIGHT		xaminee's right eye	.01 === 2.5 :	ioual a-vit	v04io04	N	0.5			
	(right)		It is exp		ı ∪.1 and 2.5. V	isual acuity below	v 0.1 is 0.1, and blindness is					
s s								1				

11 Hea	aring (left)	HEAR_LEFT	• Hearing 1 (normal), 2 (abnormal) in the	N	1	8.0
			subject's left ear			
12	hearing shifty	HEAR_RIGHT	Hearing in the examinee's right ear (normal), 2 (abnormal)	N	1	
13	Systolic Blood pressure	BP_HIGH	The heart contracts with the examinee's maximum blood pressure, and Intravascular pressure when fluid is pumped into the arteries	N/mmHg	140	
14	diastolic	BP_LWST	The patient's diastolic blood pressure at the time of completion of the heart	N/mmHg	81	
15	whole meal blood sugar (fasting blood sug	BLDS	Blood sugar (concentration of glucose in 100ml of blood) level before the examinee's meal	N/mg/dL	94	
16	total cole sterols	TOT_CHOLE	Sum of ester-type and non-ester-type (free) cholesterol in serum Normal value is 150-250 mg/dL, about 1/3 of which is non-ester-type (free) cholesterol, and the rest is cholesterol ester.	N/mg/dL	164	
17	triglyceride	TRIGLYCERIDE	Means simple lipids or neutral lipids It is a natural product made by combining glycerol with tri-molecular fatty acids as esters. It has the widest distribution among the fat derivatives that can be found in Normal value is 30–135mg/dL (0.34–1.52-mmol/ÿ) ÿ Since the triglyceride item has been added as a health checkup questionnaire since 2008, if the base year is from 2002 to 2007, the item value is treated as missing and provided.	N/ mg/dL	94	
18	HDL cholesterol	HDL_CHOLE	Cholesterol in HDL (High Density Lipoprotein) The normal value is 30-65mg/dL, a component that returns the cholesterol drawn to the cells with small particles of cholesterol to the liver and removes the bad cholesterol that has accumulated on the walls of blood vessels. ÿ Since the HDL cholesterol item has been added as a health checkup questionnaire since 2008, if the base year is from 2002 to 2007, the item value is provided as missing.	N/ mg/dL	45	
19	LDL calls leste role	LDL_CHOLE	Cholesterol in LDL (Low Density Lipoprotein) If the amount is excessively increased with very large particles of cholesterol, blood A component that accumulates on the tube wall and causes arteriosclerosis or various diseases If it is higher than 170mg/dL, it is generally regarded as hyperLDLemia. ÿ Since the LDL cholesterol item has been added as a health checkup questionnaire since 2008, if the base year is from 2002 to 2007, the item value is provided as missing.	N/ mg/dL	Example 150 Example 2130	
20 hen	noglobin	HMG	globin as a pigment protein present in blood or blood cells It is composed of (globin) and heme (heme) and serves as an oxygen carrier in the blood. g/dL 15.39843		15.3984375	
21 Urir	e Protein OLIC	_PROTE_CD	• Protein in the urine 1(-), 2(±), 3(+1), 4(+2), 5(+3), 6(+4)	N	1	
22	Serum reatinine	CREATININE	Creatinine is a dehydrated product of creatine, which is a species of endogenous anabolic Excreted by the kidneys as a malic product, its increase or decrease is related to muscle growth and movement regardless of food Serum creatinine concentration increases due to renal dysfunction Normal value 0.8~1.7mg/dL ÿ Since the serum creatinine item has been added as a health checkup item since 2008, if the base year is from 2002 to 2007, the item value is provided as missing.		0.8	
23						

	(serum	SGOT_AST	Blood test values indicating liver function, in addition to hepatocytes, heart,	N/	12	
	paper OT)	0001_A01	Blood test values indicating liver function, in addition to nepatocytes, neart, kidney, brain Enzymes that are also present in muscles, such as	mg/dL		
	AST		Concentration increases when damaged	g-uz		
			Normal value 0~40IU/L			
			A blood test level indicating liver function, ALT is an enzyme mainly present			
	(serum		in hepatocytes, and the concentration increases when hepatocytes are		45	
24	paper OT)	SGPT_ALT	damaged.	N/U/L	15	
	EVERYTHING		Normal value 0~40IU/L			
			A blood test value indicating liver function, an enzyme present in the bile duct (bile duct) in the			
	gamma paper		liver that transfers glutamic acid to peptides or amino acids to the outside. Normal value: 11-63		Example 119	
25		GAMMA_GTP	IU/L for men, 8-35 IU/L for women	N/U/L		
	teepee				Example 2114	
	smoking award		Whether the examinee is smoking			
26	and and a	SMK_STAT_TYPE_CD	1 (does not smoke), 2 (previously smoked, but quit), 3 (still smokes now)	N	1	
	womb					
	drink		Whether the examinee is under the influence			
27	Ullik	DRK_YN	of alcohol 0 (does not drink), 1 (drank)	N	1	
	wealth		N (Do not drink), Y (Drink)			
	mouth sword		Whether or not the examinee selected oral examination			
28		HCHK_OE_INSPEC_YN	For items 0	N	1	
	Whether or not	HOHK_OL_INGFLO_TN	(not tested), 1 (examined)			
	to check		N (unexamined), Y (examined)			
	Chiau		Items on whether or not the examinee has dental caries			
29	food oil	CRS_YN		N	1	
	radish		0 (none), 1 (yes)			
	/-					
			• Items 0 (none), 1 (yes) for the presence or absence of the examinee's missing teeth (tooths			
30	missing teeth	TTH_MSS_YN	that have fallen out of the dentition due to various causes (carious periodontal disease, tooth	N	1	
	The presence of almost		extraction, etc.)			
			Sudden, ste.)			
	01:					2002~
	Chiama		Items on the presence or absence of dental abrasions of the examinee			Only in
31	breast milk	ODT_TRB_YN	0 (none), 1 (yes)	N	1	
	radish					2013
	3rd generation					
	molars					
32	(I love	WSDM_DIS_YN	Is there any abnormality in the examinee's third molar (wisdom tooth)?	N	1	
	you)		0 (none), 1 (yes)			
	over					
			Whether the examinee has tartar			
33 tarta	ır	TTR_YN	0 (none), 1 (yes)	N	1	
			N (none), Y (yes)			
	data					
.	data	DATA OTD		YYYYMM	20161222	
34	date of	DATA_STDDT	Data creation date	DD	20161220	
	publication					
-	_					

ÿ Ir비고란에 표시가 되어 있는 개방항목의 경우 ame of the individual code value is provided, the server load and download speed decrease due to the increase in data amount

Only the code value is provided due to a problem such as:

- ÿ Since administrative districts are each coded by year, changes should be thoroughly reviewed and utilized. For example, in the case of Sejong Special Self-Governing City, as it was newly incorporated from 2012, there is no corresponding item in the data up to 2011.
- ÿ The health checkup information is the main results and questionnaire response data of the 1st general health checkup and life-transition health checkup (2008–) for the examinee in the relevant year, and data on the general health checkup of medical benefit recipients are not included.
- ÿ Since major examination items were changed due to the reorganization of the health examination system in 2009, changes such as yearly health examination system, health examination items, questionnaire items and types, and variable names should be thoroughly reviewed and utilized.

구분	2002-2007년	2008년	2009-2015년
kidney			
weight			
Waist circumference			
systolic blood pressure			
diastolic blood pressure			
Pre-meal blood sugar (fasting blood sugar)			
total cholesterol			
triglycerides			
(triglyceride) HDL cholesterol			
LDL cholesterol			
hemoglobin			
urine protein			
serum creatinine			
(Serum Geotee) AST			
(Serum Gioti) ALT			
Gamma GT			

Section 4 How to download country-focused open data

By accessing the public data open service (http://nhiss.nhis.or.kr/op/it/index.do), you can conveniently download national-focused open data.

Introduction of how to use public data open services

1) 단계 Access the website for opening country-focused open data and select the "Download" menu on the first screen



공공데이터 개방서비스 첫 화면에서 다운로드 메뉴 선택

2) Lallect the information and year you want to download from among medical history, drug prescription, and health checkup information



3) **Etal**: nter brief user information such as age group, gender, user residence (province and province), purpose of use, etc. (The main user base and purpose of use of open data users are collected and reflected in the establishment of policies related to data provision in the future. want to)



4) **단계** Go to the final download screen and download the attached file. The attached file includes the user's manual and data for the year selected by the user.





ÿ Provide national-focused open data utilization guides and examples

Through the website for opening national-focused open data, frequently asked questions from users are organized and provided as a guide for use. In addition, by providing various domestic and foreign cases developed using national-focused open data, we support the effective use of data by the people who want to utilize national-focused open data.

Support for online inquiry function using the website for opening national-focused open data

You can make an online inquiry related to the use of country-focused open data through the 'Inquiry' function on the website for opening national-focused open data. In addition, if there are any suggestions or errors related to data use, a function to report it to the manager is provided to enable effective communication with users.

Support for providing demo data before downloading country-focused open data

By allowing the user to check the demo data of the relevant information before downloading the country-focused open data, the user can User convenience is increased as it can be checked in advance whether information is included in the country-focused open data.



- ÿ Do I have to go through a separate application and examination process to utilize the nationally-focused open data?
- ÿ No. According to the user manual, you can access the National Health Insurance Data Sharing Service (NHISS) web page and download it freely.

 There is no need to go through the application and review process.
- ÿ Can it be used for commercial purposes, such as opening medical service-related applications or sites by utilizing national-focused open data?
- ÿ It is possible. State-focused open data is public data that can be freely used by the private sector and is not in conflict with individual laws (medical law, etc.)

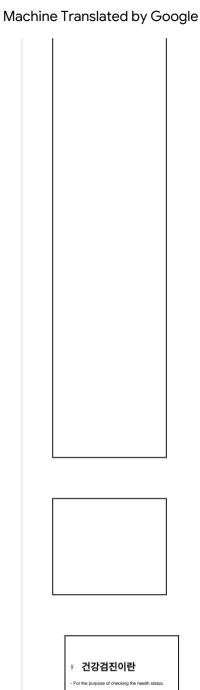
 It can be used for various purposes within the above.
- \ddot{y} I am curious about the usable range of open country-focused data. Can it be used for academic and research purposes?
- ÿ In order to be used for academic and research purposes, health-related data on a large number of samples must be provided. However, country-focused open data is available through the web.

 It is impossible to provide a large amount of data because the information is provided. For this reason, even in the construction process, partial editing or processing of data was performed. In addition, since the data is provided to the general public, it is only provided to academic researchers after being reviewed.

The number of open items is smaller than that of data. In conclusion, I think it would be more effective to use the sample research DB that the National Health Insurance Corporation has built and provided for academic and research purposes.

9 고유식별정보 to uniquely identify an individual in accordance with the law 1) Resident, etc. Lock number ŷ Passport number ŷ Luck License number of previous driver's license ŷ Refers to allien registration number (Personal Information Protection Act Enforcement Decree Article 19 (Scope of Uniquely Identifiable Information))

Machine Translated by G	oogle
ÿ For detailed information on what kind of health checkup results are provided, refer to ÿSection 3 National Health Information Data Details – 4 Health Checkup Information- a. concept and phr. Gender-1) Refer to the concept	42 42
]



For the purpose of checking the health status, preventing and early detection of diseases, medical examinations such as medical examinations, physical examinations, diagnostic examinations, pathological examinations, and radiology examinations are conducted through a health examination institution.

fealth Examination Framework Act Article 3 (Definition

No. 1)

🤋 국가건강검진이란

State and local governments

health check-up

 \ddot{y} Health checkup for infants and toddlers in accordance with the Mother and Child Health Act

Jin

 \ddot{y} Health checkup for infants and toddlers in accordance with ÿInfant Care Actÿ

 \bar{y} Cases of elementary, middle and high school students in accordance with the \bar{y} School Health Act

Intensive

examination ÿ Youth health checkup in accordance with ÿYouth Welfare Support Actÿ

- Health checkup under ÿNational Health
Insurance Actÿ - General health checkup under

ÿOccupational Safety and Health Actÿ

 \ddot{y} Health checkup in accordance with $\ddot{y}\text{Medical}$

Benefit Act

ÿ Cancer screening according to the ÿCancer

Control Act

ÿ Health checkup in accordance with ÿElderly Welfare Actÿ -

Other health checkups prescribed by Ordinance of the Ministry of Health and Welfare

요단백 판정 기준

If protein is excreted more than 500 mg per day for adults and 4 mg or more per square area of body surface area for 1 hour in children, evident proteinuria. dipstick method) to the degree of discoloration to green within 60 seconds after soaking urine in the test paper. Silver is +3, 1000mg/dL is judged as +4

🤊 시력 측정 방식

The function of the eye to distinguish the shape or existence of an object is called the angle of view, and the degree of visual acuity is to measure the distance between the first two points that can be distinguished from two points as two points, and this is displayed as a visual and the reciprocal expressed as a quantity proportional to

🤊 청력이란

The decibel of the auditory sense to the intensity of sound.

Displayed in (dB) 0

dB for normal people, the number increases with hearing loss

