Obtaining phenotype and outcome data from health records:
China Kadoorie Biobank experience

Zhengming CHEN
CKB Principal Investigator
Professor of Epidemiology
Nuffield Dept. of Population Health
University of Oxford, UK
(zhengming.chen@ndph.ox.ac.uk)

International Cohorts Summit, Duke University, USA,
26-27 March 2018

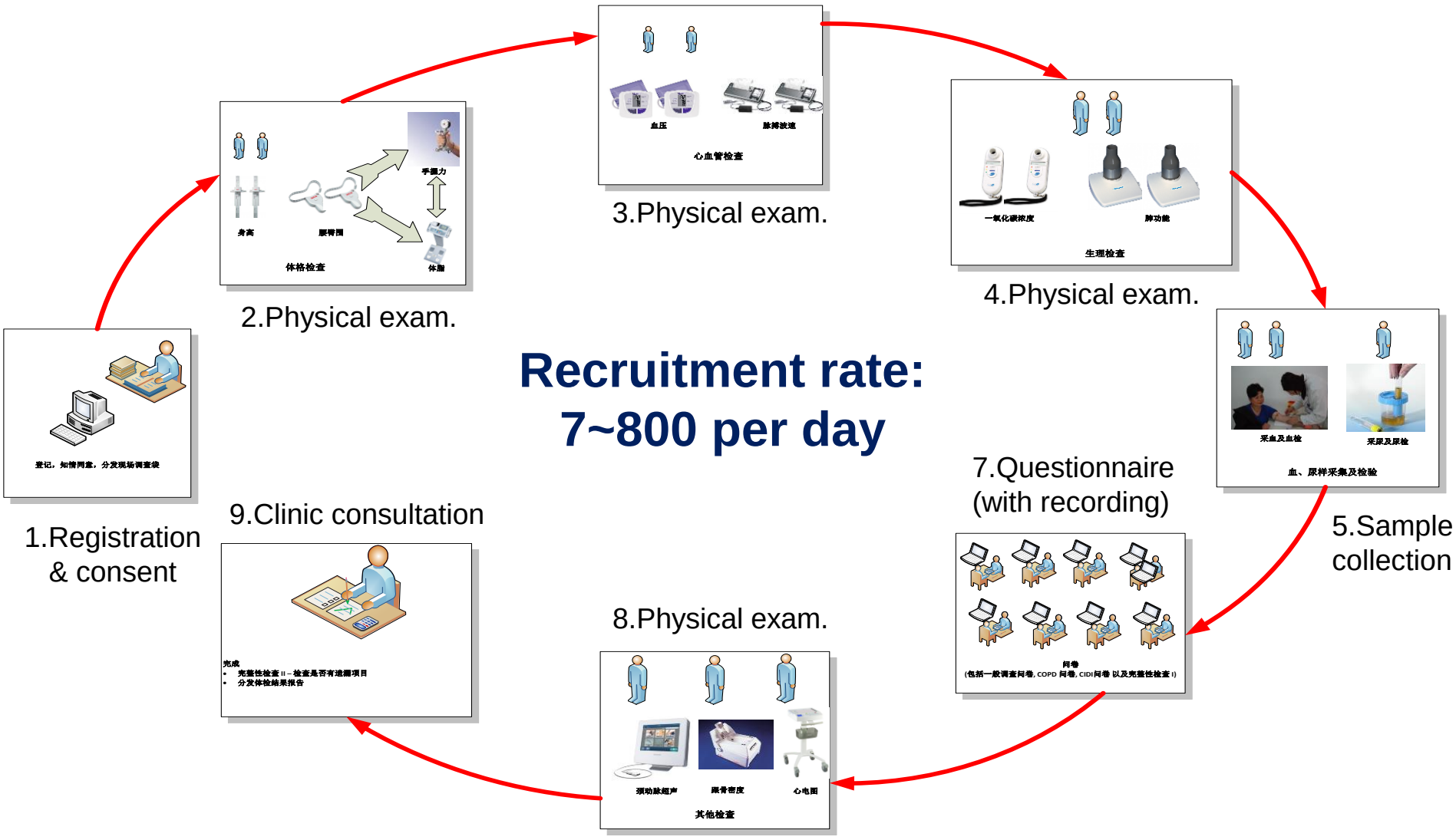
China Kadoorie Biobank (CKB)

- >512K recruited from 10 localities in 2004-08
- Participants interviewed, measured, and gave plasma and DNA (urine) for long-term storage
- All followed up indefinitely via electronic record linkage to deaths and ALL hospital episodes
- Periodic resurvey of 5% surviving participants (for enhancements and sources of variation)

Informed consent for linkages to health records and unspecified research use of stored samples

CHINA KADOORIE BIOBANK
中国慢性病前瞻性研究

CKB: Clinical stations at local assessment centre



The clinic visit took 60-90 minutes, with daily statistical monitoring

CKB: Supported by >90 bespoke IT systems

Asset and Research Management

CDAS
PvdHospitals
ConsentFormTracking
ResearchTracker
Geocoder
Standardiser
Labeliser
SpouseChecker
MMS
StudyIDAuth

Long-Term Follow Up and Validation

CRDWeb
OutcomePVD
iCASE
OV-Manage
LTFollowUp
OV-Reports
PvdManager
OV-Viewer

Interface to External Systems

BgiImportExportJob
DeviceDataConverter
ExternalShipments
LabResultFetcher

Monitoring and Reporting

FTP
LatestActionLog
LogViewer
Reporting
SendLogJob

A Typical Range of Networked Computers Distributed Across CKB Study Centres in the UK and China.

U.K. CKB International Coordination Centre

Secure Communication

China CKB National Coordination Centre

CKB Field Clinics

Regional Centres

Labs

AddRC
CAS
ClinicalLocations
PassMan
UserMan

Centre and User Management

BloodTest
PulseTrace
BodyComposition
QcSelect
DrList
ResurveySelect
GadgetManager
ScanMan
NIDCheck
Sinput
NonResponderSurvey
Sinspire
ParticipantDetails
Sinvite

Field Data Collection

CKB Website
CKB Wiki
KSCDC Website
Sinmail
Teleport

Communication Tools

Babel
NodeStatus
DbBackup
Sinit
DBTourist
Sinserver
DirSync
Sinsync
KeyServer
Sintegrator2
LaptopActivator
UpdateMan

IT Systems Management

BarCodeChecker
NccStore
BoxHistory
Noted
BrowsePacks
SampleAli
CryovialHandler
SampleLogging
DNAConcentrator
SampleMonitor
IccStore
SampleTracking
NccSampleLocator
StorageAreaConfig

Bio-Sample Management

CHINA KADOORIE BIOBANK

CKB: Fully established with 10-year follow-up

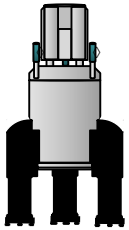
Questionnaire
SES, smoking, alcohol, tea, diet, physical activity, indoor air pollution, sleep, reproductive patterns, medical history

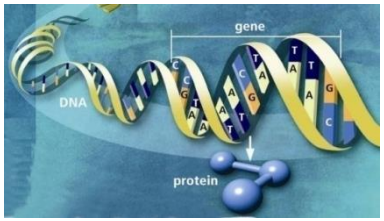
Measurements
Blood pressure, height, weight, lung function, heart rate, bone density, exhaled CO, ECG, cIMT, ambient temperature, ambient air pollution, blood lipids, metabolites, proteomics, infectious markers, genetics

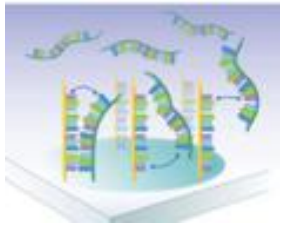
Electronic health records
>1,300 different diseases, 43K deaths, <5K lost to follow-up, ~0.9 million hospitalizations, >100 million chargeable items

www.ckbiobank.org

Data is growing rapidly







CHINA KADOORIE BIOBANK
中国慢性病前瞻性研究

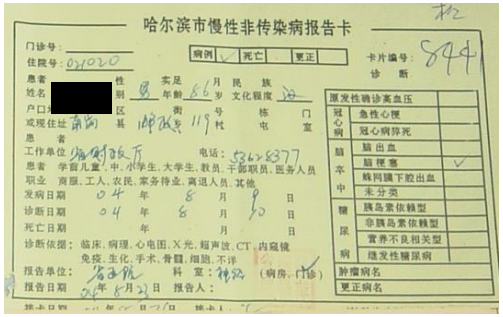
CKB: Follow-up through record linkages

Death registries

Active follow-up

Outcome Follow up in CKB

Disease registries



National health insurance

CHINA KADOORIE BIOBANK
中国慢性病前瞻性研究

National health insurance system in China

(supplementing death and cancer registries)

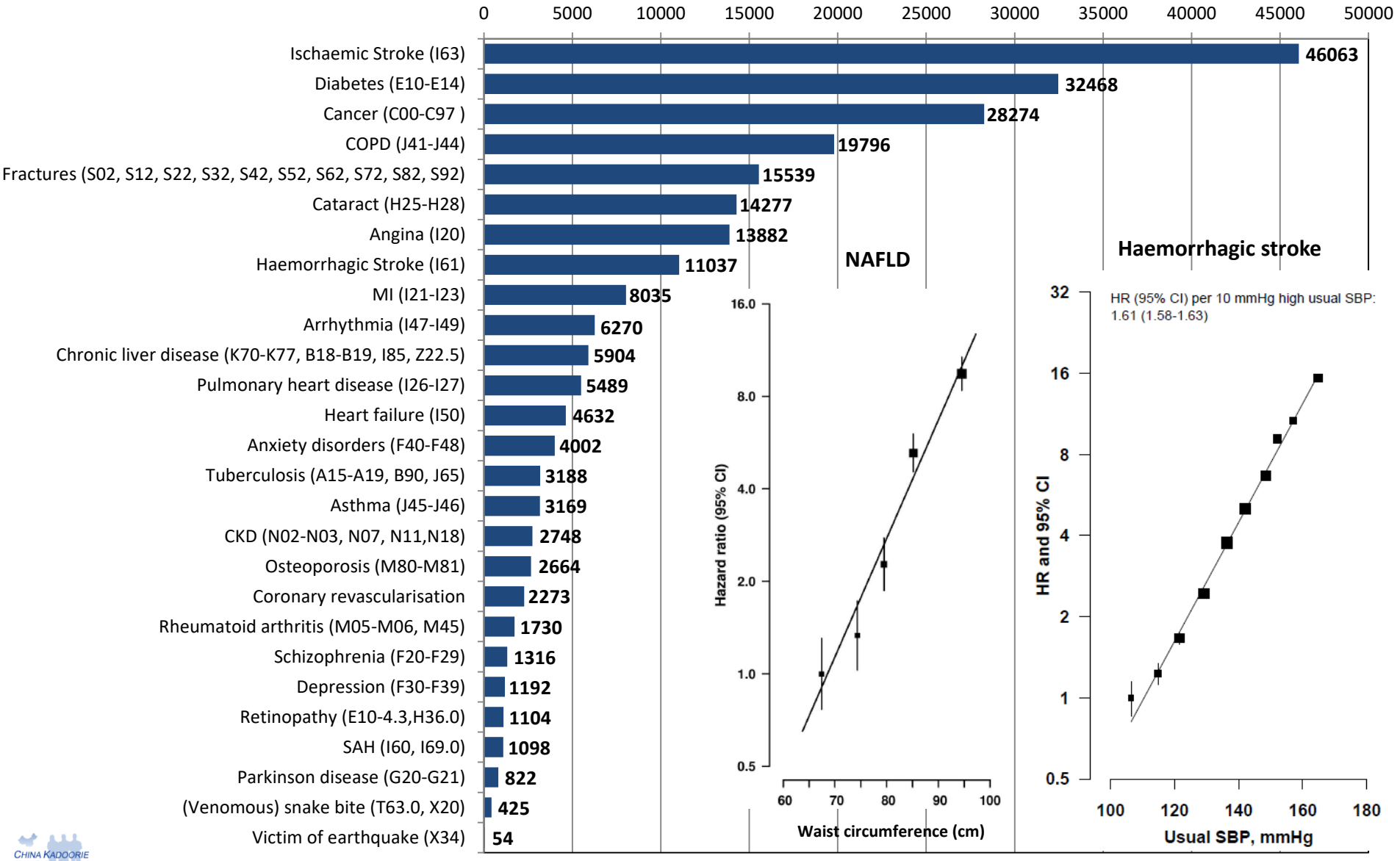
- Introduced during 2004-6, with almost universal coverage in CKB areas by 2010
- Multiple disease diagnoses, with ICD-10 codes plus disease descriptions and >2,500 procedure codes
- Managed electronically at city level, with detailed chargeable items for reimbursement purposes
- Lacks certain details (e.g. cancer pathology) required for disease sub-phenotyping

Nearly all CKB participants now linked to the health insurance databases via unique national ID number

CHINA KADOORIE BIOBANK

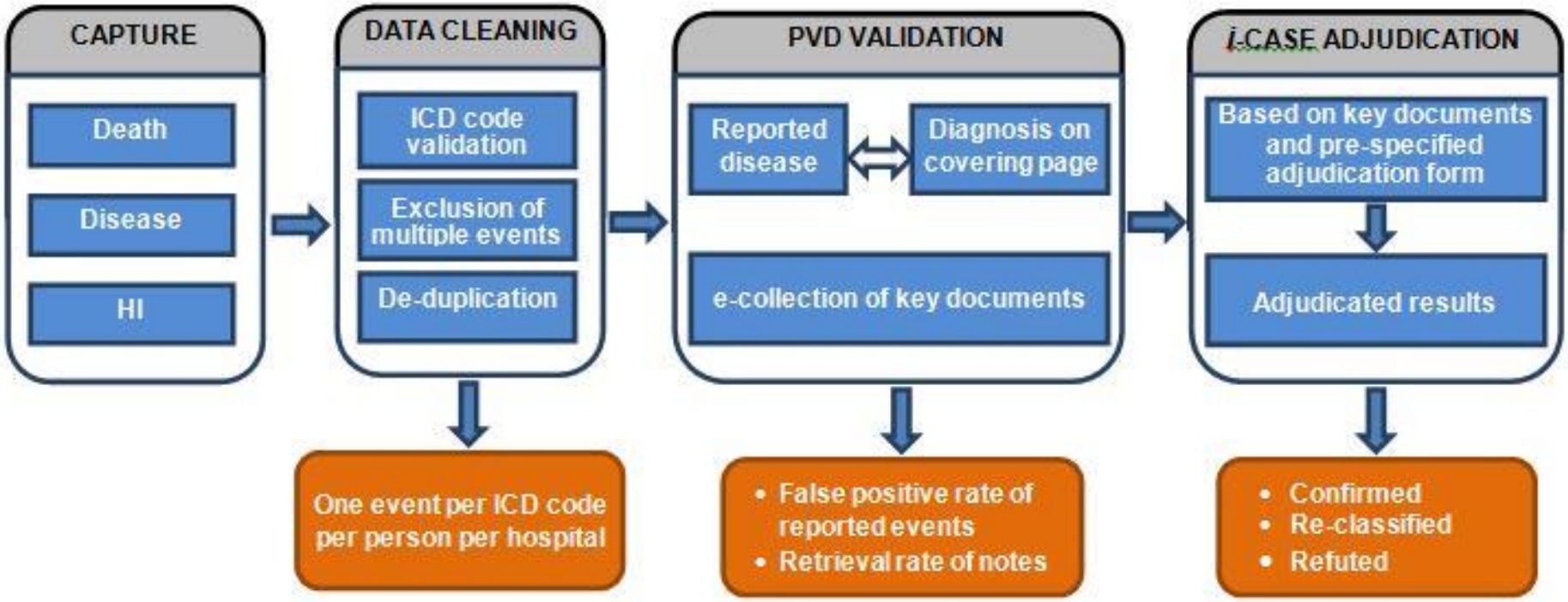
CKB: Participants with selected diseases in 10 years

(43K deaths, 0.9M hospital admissions; 2017 HI data are being processed)



CHINA KADOORIE BIOBANK
中国慢性病前瞻性研究

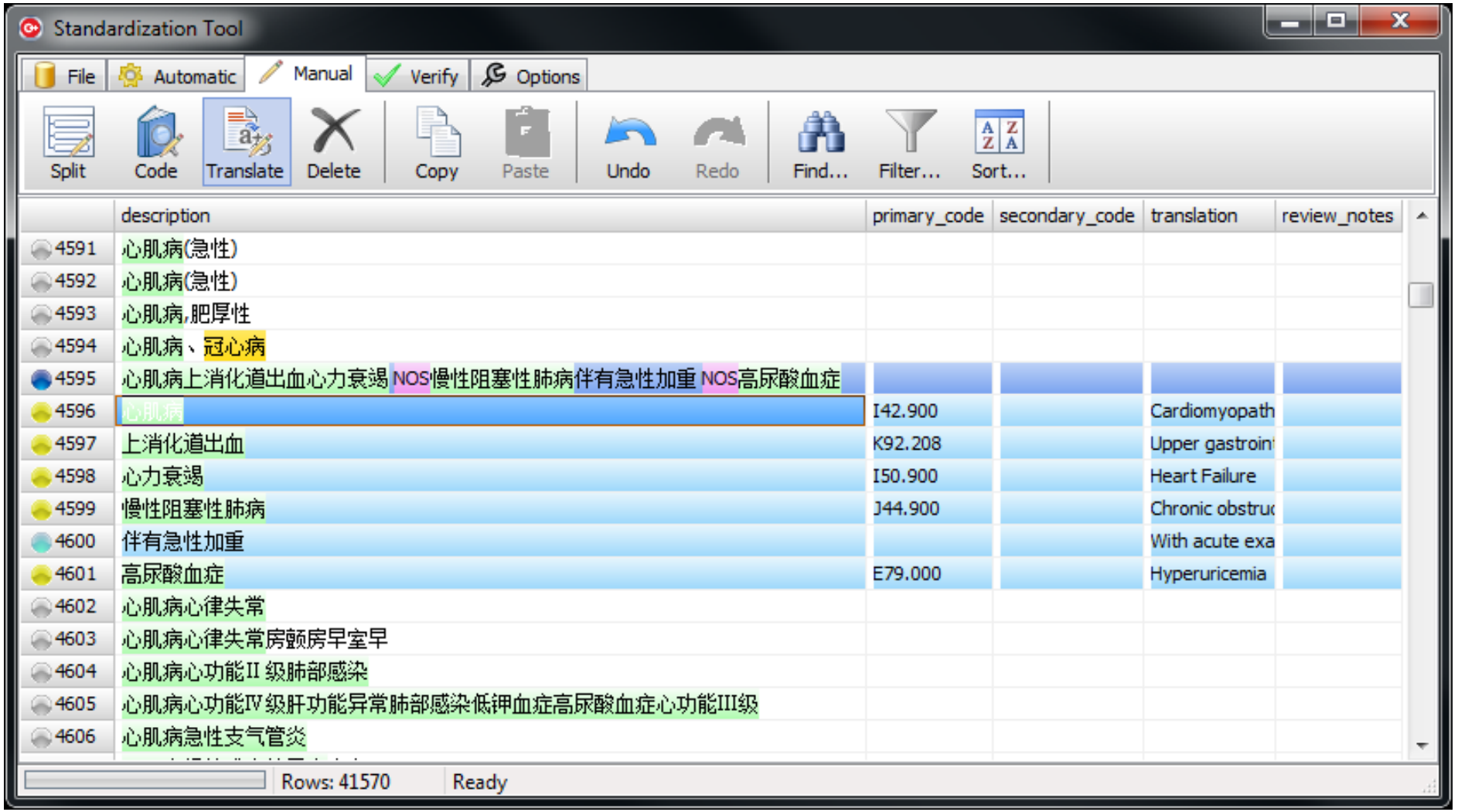
CKB: Procedures for improving disease phenotyping



Pilot study of ~1000 cases for specific disease before deciding whether to undertake systematic adjudication

CHINA KADOORIE BIOBANK
中国慢性病前瞻性研究

CKB: Disease standardisation and coding tool



CHINA KADOORIE BIOBANK

CKB: Verifying reported diagnosis

Overview of Case Validation

Personal Information	
Name	smith
Sex	Male
Date of birth	1952-04-03
Hospitalisation number	12

Hospital Admissions

Admission date	Discharge date	Is this an electronic note?
2008-05-05	2008-05-15	Yes
2010-01-04	2010-01-07	No

Vital Status	
What is the patient's vital status at discharge from the last note based on the discharge date in this hospital?	Alive

Main Diagnosis	
Please choose the disease sub-type from the	Ischemic stroke

The following section shows the updated information for this patient. Please check it carefully to confirm that the details are all correct.

Is this information correct?

(•) Yes ( ) No

Please enter your user name and password and select 'Finish' to sign-off the validation. After this you will not be able to make any changes.

User name

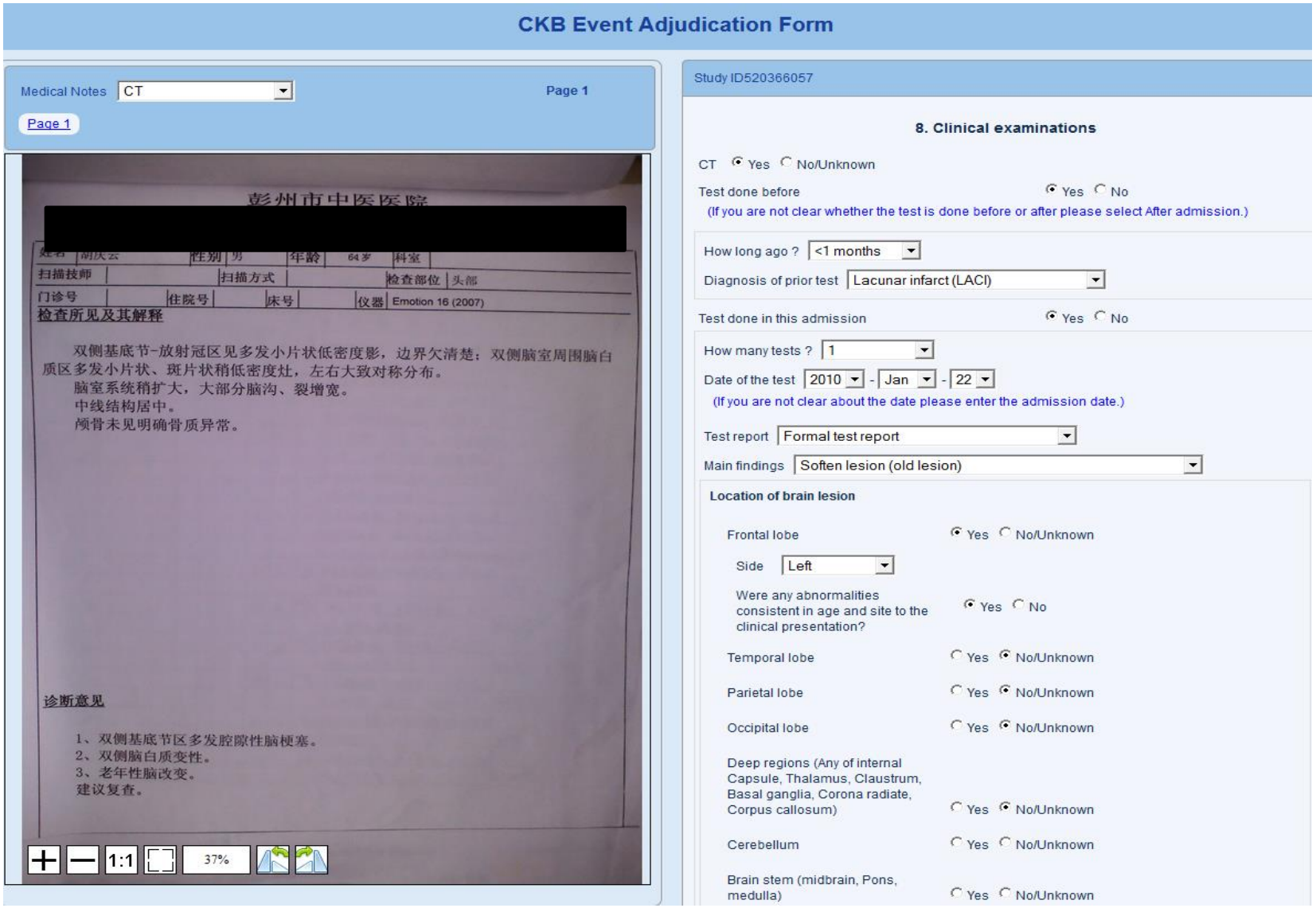
Password

Finish

Previous

CKB: Adjudicating & phenotyping major diseases

(>70K adjudicated: 30K stroke, 25K IHD, 15K cancer, >3K CKD)



CHINA KADOORIE BIOBANK
中国慢性病前瞻性研究

CKB: "traffic" light approach for outcome data

Clipboard | Font | Alignment | Number | Styles | Cells

B15 — Unknown and unspecified causes of morbidity (R69)

	A	B	C	D
1	icd10	desc	event_count	participant_count	pvd_
3	I25	Chronic ischaemic heart disease (I25)	141900	46187	
4	I63	Cerebral infarction (I63)	164372	46014	
5	K52	Other noninfective gastroenteritis and colitis (K52)	105305	40866	
6			51989	36064	
7	K29	Gastritis and duodenitis (K29)	62053	32990	
8	I67	Other cerebrovascular diseases (I67)	41861	22621	
9	E14	Unspecified diabetes mellitus (E14)	91025	20778	
10	G45	Transient cerebral ischaemic attacks and related syndromes (G45)	27600	18799	
11	E11	Non-insulin-dependent diabetes mellitus (E11)	34733	18773	
12	J98	Other respiratory disorders (J98)	27109	16926	
13	J00	Acute nasopharyngitis [common cold] (J00)	33758	16629	
14	M54	Dorsalgia (M54)	27798	16589	
15	R69	Unknown and unspecified causes of morbidity (R69)	32567	16436	
16	M51	Other intervertebral disc disorders (M51)	26480	15944	
17	J18	Pneumonia, organism unspecified (J18)	24013	15592	
18	J40	Bronchitis, not specified as acute or chronic (J40)	32628	14232	
19	J44	Other chronic obstructive pulmonary disease (J44)	45249	14018	
20	M47	Spondylosis (M47)	23142	13968	
21	I20	Angina pectoris (I20)	27723	13755	
22	R42	Dizziness and giddiness (R42)	23106	12023	
23	M13	Other arthritis (M13)	20172	11631	
24	T14	Injury of unspecified body region (T14)	17998	11595	
25	J11	Influenza, virus not identified (J11)	21770	11454	
26	R53	Malaise and fatigue (R53)	22566	11300	
27	I61	Intracerebral haemorrhage (I61)	42288	10887	
28	N20	Calculus of kidney and ureter (N20)	18386	10777	
29	H26	Other cataract (H26)	14797	9927	
30	K81	Cholecystitis (K81)	14869	9769	
31	K80	Cholelithiasis (K80)	15580	9685	
32	J42	Unspecified chronic bronchitis (J42)	17705	9536	
33	J20	Acute bronchitis (J20)	14208	9339	
34	R10	Abdominal and pelvic pain (R10)	13600	9285	
35	K30	Dyspepsia (K30)	11704	8804	
36	I21	Acute myocardial infarction (I21)	25137	7865	
37	J02	Acute pharyngitis (J02)	14508	7824	
38	K05	Gingivitis and periodontal diseases (K05)	12304	7265	
39	I84	Haemorrhoids (I84)	9495	7193	
40	N76	Other inflammation of vagina and vulva (N76)	17261	7132	
41	I69	Sequelae of cerebrovascular disease (I69)	19670	6762	

Future work for disease phenotyping

- Standardising and ICD-10 coding new events collected
- Processing and incorporating >100M chargeable items data to enhance disease phenotyping
- Extending outcome adjudication to several other diseases (e.g. heart failure, chronic liver disease)
- Developing automated algorithm to sub-phenotype stroke and other diseases according to clinical criteria
- Piloting collection of discharge summary pages and tumour tissue samples

CKB: Open data access platform

(www.ckbiobank.org)

嘉道理 KADOORIE

MRC Clinical Trial Service Unit

British Heart Foundation

CANCER RESEARCH UK

NSFC

wellcome trust

CHINA KADOORIE BIOBANK 中国慢性病前瞻性研究