

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



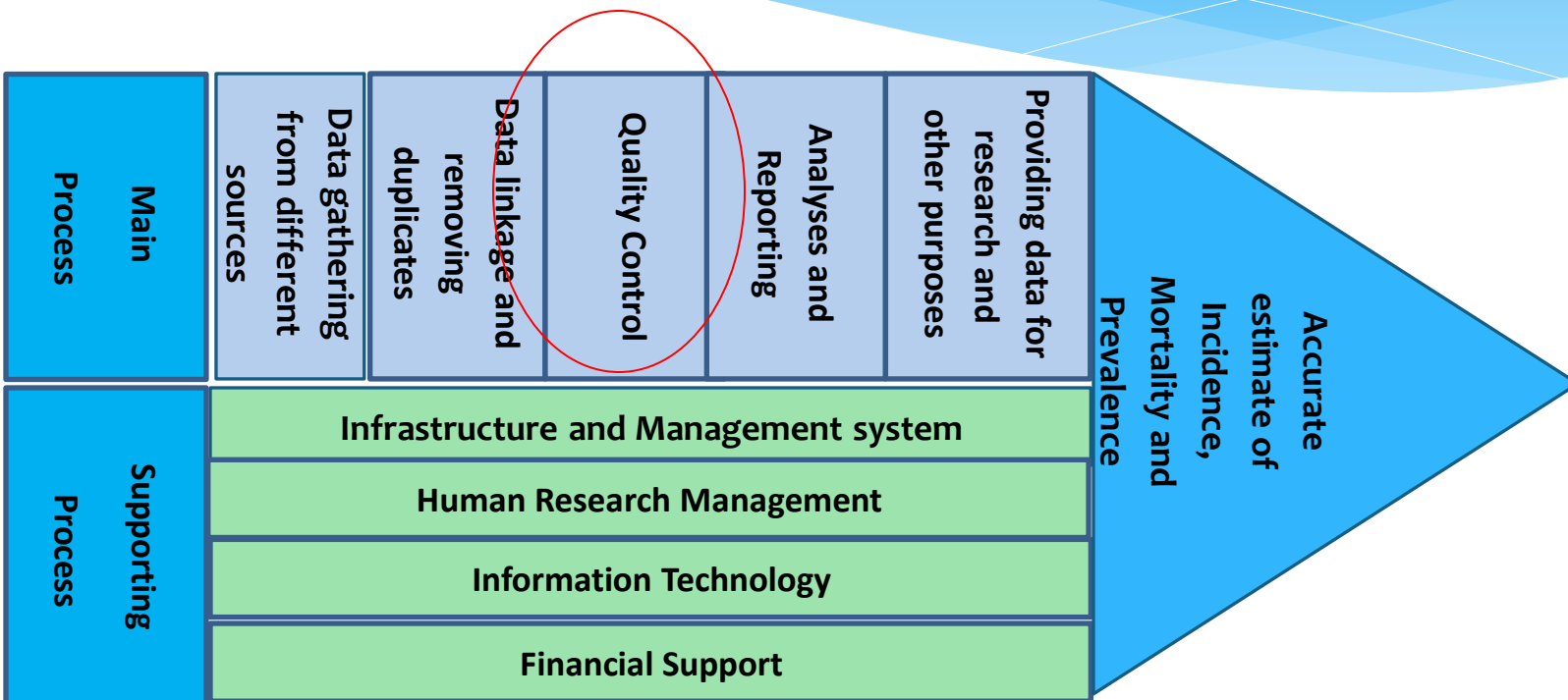
کارگاه ثبت بیماری ها

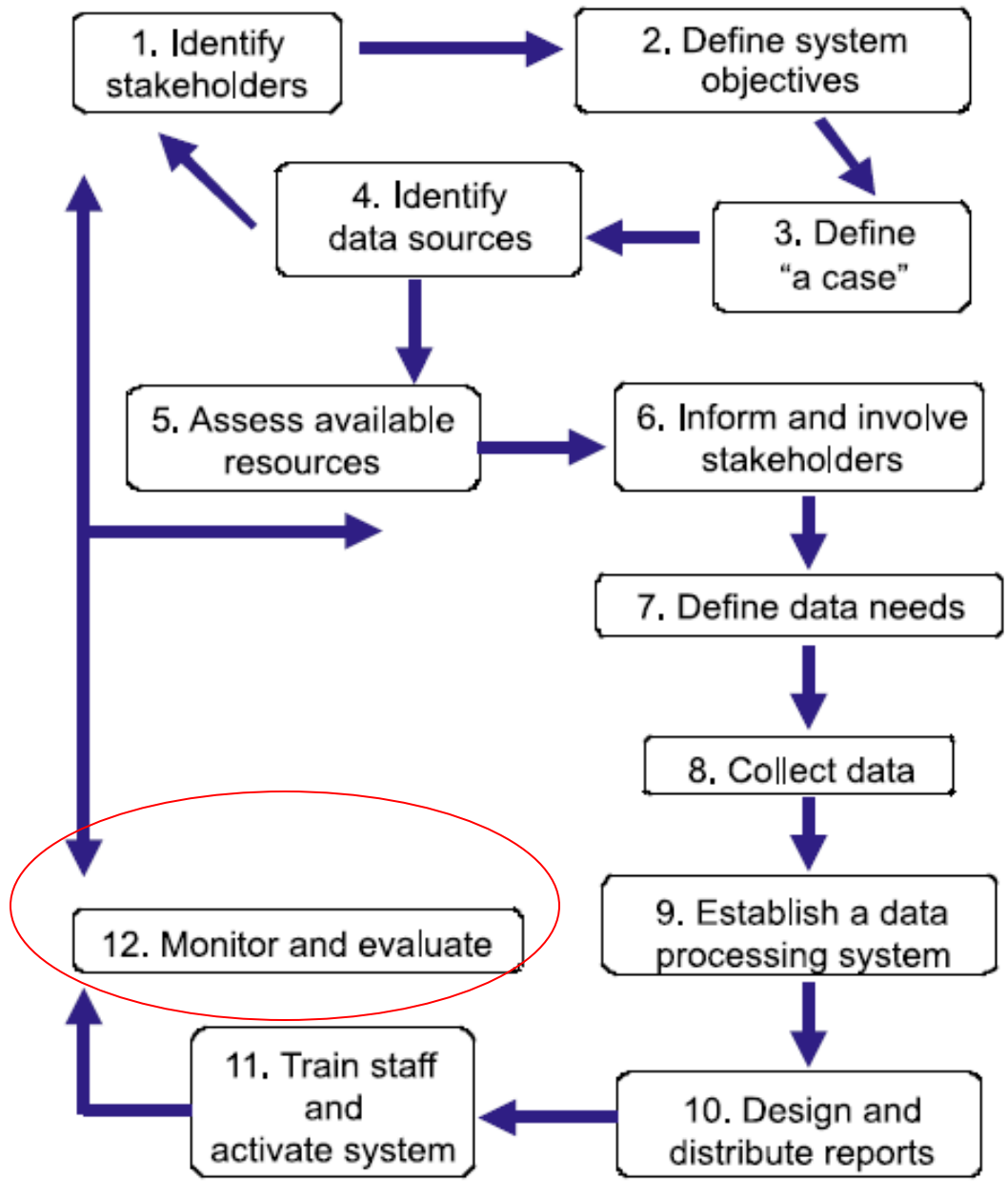
کیفیت داده در رجیستری

دکتر عباس شیخ طاهری

دکترای مدیریت اطلاعات سلامت
عضو هیات علمی و سرپرست ثبت بیماری ها
دانشگاه علوم پزشکی ایران

Processes





کیفیت داده رجیستری

* کیفیت بر مبنای ویژگی‌های از پیش مشخص تعریف می‌شود.

* روش بر اساس هدف ثبت، نوع داده، روش گردآوری، منابع داده و برنامه زمانی برنامه ثبت، میزان بودجه فرق دارد.

* روش‌ها از پیش مشخص باشد.

Quality attributes (characteristics)

Accuracy (صائب / صحیح بودن)

* راست و درست بودن

* توافق میان ارزش داده و منبعی که فرض بر درستی آن است

Timeliness بموقع بودن / بهنگامی

* محدوده زمانی که برای ثبت داده مجاز است.

* محدوده زمانی برای گزارش اطلاعات

completeness کامل بودن / جامعیت

* وجود تمام داده‌های لازم

* ضرورت کامل شدن تمام عناصر داده

Relevancy

* معنی دار بودن داده برای کاربردی که به خاطر آن گردآوری شده‌اند.

* ارتباط معنی دار بین داده‌ها و کاربرد مورد نظر

Consistency (همسانی)

* عدم تناقض بین داده ها

* سازگاری درونی بین داده ها

* در مورد صفات مشابه انتظار این است که داده ها مشابه باشند (مثلا سن)

* در مورد صفات متفاوت ولی مرتبط انتظار می رود داده ها متناقض نباشند (نوع اقدام و جنسیت بیمار)

Definition (تعریف شدگی)

* وجود معانی مشخص برای داده‌ها و عناصر داده به منظور فهم یکسان آنها

* داده‌ها و عناصر اطلاعاتی و مقادیر باید دارای تعاریف واضح باشند تا کاربران فعلی و آتی معنی داده را بدانند.

Precision

- * تعیین ارزش‌های قابل قبول یا طیفی از ارزش‌ها به اندازه‌ای که کاربرد مورد نظر را پشتیبانی کند
- * مقادیر داده باید به اندازه کافی دقیق باشد تا از آن کاربرد یا فرایند خاص حمایت کند.
- * به‌عنوان نمونه، گرد کردن مقدار دوز دارویی که بر حسب "میلی‌گرم" تجویز شده است به نزدیکترین "گرم" به هیچ‌عنوان کار درستی نیست.

Reliability

پایایی / قابل اعتماد بودن

* **نتایج یکسان** با تکرار عملیات مختلف گردآوری و تحلیل داده ها

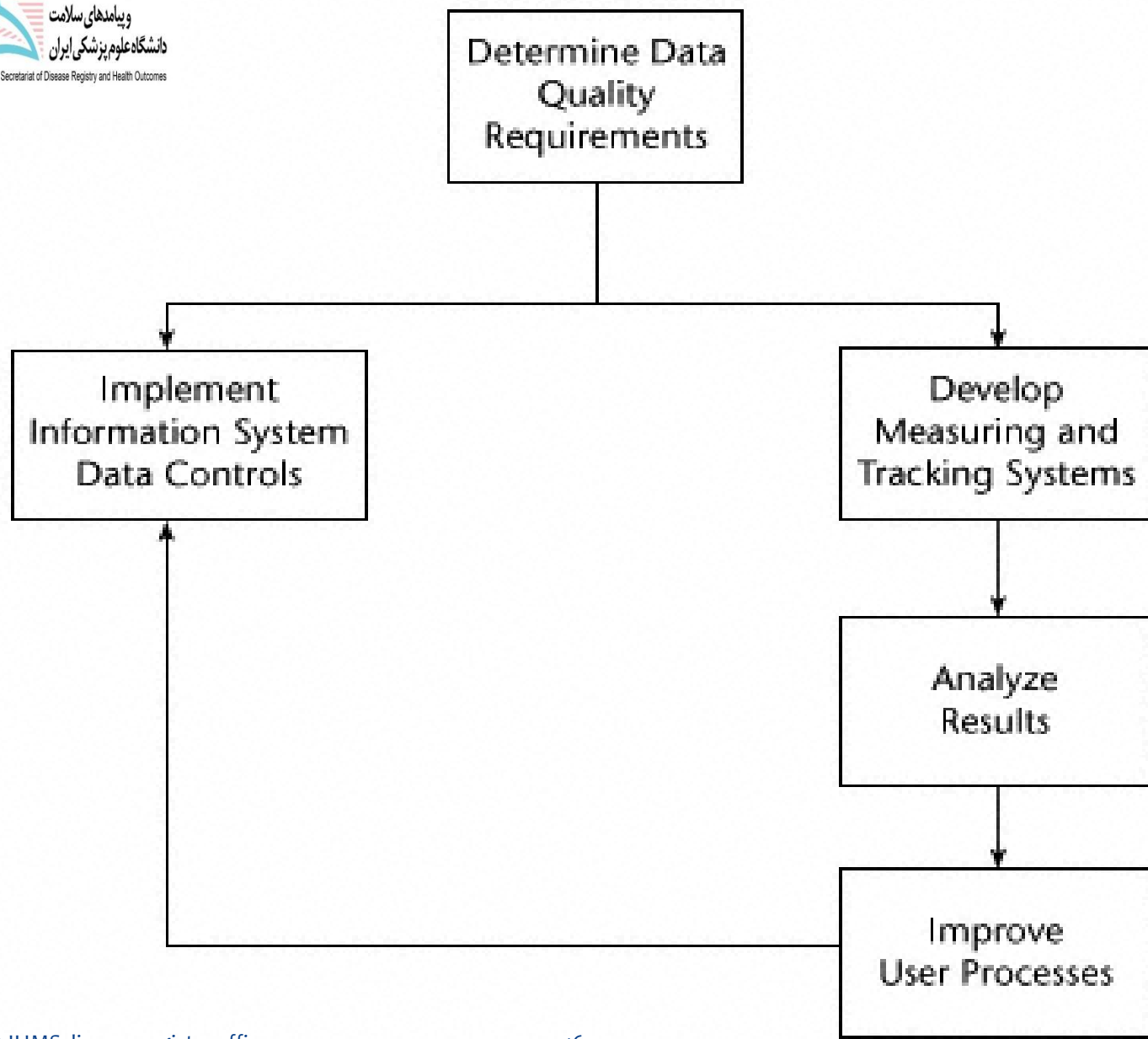
Comparability قابل قیاس بودن

* قابل مقایسه بودن داده ها

* اگر داده ها بر اساس یک تعریف مشترک نباشند نمی توان آنها را مقایسه کرد.

Coverage پوشش

* داده ها باید انعکاس دهنده کافی جامعه تحت پوشش باشد



	<i>Pr. Demographic Data</i>	<i>Chief Complaint</i>	<i>Lab data</i>	<i>Diagnostic code</i>	<i>STATISTICS</i>	<i>ETC</i>
Accuracy	H	L	H	?	?	
Completeness	H	L	L	?	?	
Consistency	H	L	H	?	?	
Currency	H	L	L	?	?	
Timeliness	H	L	M	?	?	
ETC						



Review

Evaluation of data quality in the cancer registry: Principles and methods. Part I: Comparability, validity and timeliness

Freddie Bray^{a,b,*}, D. Max Parkin^{c,d}

^aDepartment of Clinical and Registry-based Research, Cancer Registry of Norway, Institute of Population-based Cancer Research, Montebello, N-310 Oslo, Norway

^bDepartment of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Norway

^cClinical Trials Service Unit and Epidemiological Studies Unit, University of Oxford, Oxford, UK

^dCancer Research UK, Centre for Epidemiology, Mathematics and Statistics, London, UK

ARTICLE INFO

Article history:

Received 1 September 2008

Received in revised form

5 November 2008

Accepted 10 November 2008

Available online 29 December 2008

Keywords:

Cancer registry

Data quality

Comparability

Completeness

Validity

ABSTRACT

The value of the modern cancer registry and its ability to carry out cancer control activities rely heavily on the underlying quality of its data and the quality control procedures in place. This two-part review provides an update of the practical aspects and techniques for addressing data quality at the cancer registry. This first installment of the review examines the factors influencing three of the four key aspects, namely, the comparability, validity and timeliness of cancer registry data. Comparability of cancer data may be established through a comprehensive review of the registration routines in place. Validity is examined via numerical indices of that permit comparisons with other registries, or, within a registry, over time, or with respect to specified subsets of cases. There are no international guidelines for timeliness at present, although specific standards for the abstraction and reporting of registry have been set out by certain organisations.

Review

Evaluation of data quality in the cancer registry: Principles and methods Part II. Completeness

D. Max Parkin^{a,b,*}, Freddie Bray^{c,d}

^aClinical Trials Service Unit and Epidemiological Studies Unit, University of Oxford, Oxford OX3 7LF, UK

^bCancer Research UK, Centre for Epidemiology, Mathematics and Statistics, London, UK

^cDepartment of Clinical and Registry-based Research, The Cancer Registry of Norway, Institute of Population-based Cancer Research, Montebello, N-310 Oslo, Norway

^dDepartment of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Norway

ARTICLE INFO

Article history:

Received 1 September 2008

Received in revised form

5 November 2008

Accepted 10 November 2008

Available online 6 January 2009

Keywords:

Cancer registry

Data quality

Completeness

Quality control

ABSTRACT

The completeness of cancer registry data – the extent to which all of the incident cancers occurring in the population are included in the registry database – is an extremely important attribute of a cancer registry. Only a high degree of completeness in case-finding procedures will ensure cancer incidence rates and survival proportions are close to their true value. This second instalment of a two-part review of data quality methods at the cancer registry, focuses on the principles and techniques available for estimating completeness, separating methods into those that are semi-quantitative – in that they give an indication of the degree of completeness relative to other registries or over time, and more quantitative techniques – those that provide a numerical evaluation of the extent to which all eligible cases have been registered.

- * Comparability
- * Completeness
- * Validity or accuracy
- * Timelines

قابل قیاس بودن

- * رعایت کامل تعاریف حداقل داده‌ها
- * تدوین راهنمای ثبت کاملاً متناسب با راهنمای ملی و بین‌المللی
- * ارزیابی برنامه‌های آموزشی در زمینه موردیابی و گردآوری داده‌ها برای تمام کارشناسان
- * بکارگیری استانداردهای بین‌المللی برای طبقه‌بندی داده
- * استفاده از استانداردهای تعریف شده برای تعاریف

تاریخ بروز؟

تعریف: منظور تاریخ بروز (اولین تشخیص) سرطان مشول ثبت است.

نکته: طبق پیشنهاد IARC ضرورت ندارد که این تاریخ دقیقاً تاریخ اولین تشخیص باشد زیرا تعیین این تاریخ بسیار سخت است (۱۱). در صورتی که تاریخ اولین تشخیص بیماری به‌طور قطع مشخص نیاتند، معیارهای زیر (به ترتیب اولویت) برای تعیین تاریخ تشخیص سرطان (تاریخ بروز) مد نظر قرار خواهد گرفت:

- تاریخ اولین تایید هیستولوژی و سیتولوژی آن تومور: در خصوص این تاریخ، به ترتیب موارد زیر مد نظر قرار می‌گیرد:
 - ۱) تاریخ نمونه برداری مدنظر قرار گیرد
 - ۲) در صورت عدم دسترسی به تاریخ نمونه‌برداری، تاریخ دریافت نمونه در آزمایشگاه مدنظر قرار گیرد
 - ۳) در صورت عدم دسترسی به تاریخ دریافت نمونه، تاریخ گزارش پاتولوژی مدنظر قرار گیرد.در صورت عدم دسترسی به اولین تاریخ تایید هیستولوژی و سیتولوژی، به‌ترتیب تاریخ‌های زیر مدنظر قرار می‌گیرد:
- تاریخ اولین پذیرش در بیمارستان به دلیل آن تومور
- در مورد بیمارانی که تنها ویزیت سرپایی داشته‌اند، تاریخ اولین ویزیت به دلیل آن تومور مدنظر قرار گیرد.
- سایر تاریخ‌های ذکر شده در مدارک موجود به‌عنوان تاریخ تشخیص
- تاریخ فوت بیمار در صورتی که اطلاعاتی از سایر منابع در دست نیاتند و بیمار به دلیل بدخیمی فوت کرده باشد
- تاریخ فوت بیمار در صورتی که بدخیمی در اتوپسی مشخص شده باشد
- در هر صورت، تنها یک تاریخ بر اساس سلسله مراتب فوق باید به‌عنوان تاریخ تشخیص انتخاب شود و تاریخ انتخاب شده، نباید بعد از تاریخ شروع درمان، قطع درمان یا فوت بیمار باشد (۱۲).

نکته: ثبت تاریخ الزامی است.

اعتبار داده ها

- * نسبتی از موارد ثبت شده که مقدار آنها دقیقاً مقدار واقعی باشد
- * وابسته به دقت منبع داده، تخصص افراد در گردآوری، موردیابی و کدگذاری است

اعتبار داده ها

basis of diagnosis *

- * هیستولوژی تومور اولیه
- * هیستولوژی تومور ثانویه
- * بررسی بالینی (تصویربرداری)
- * تشخیص بالینی (قضاوت بالینی)
- * گواهی فوت

- * **histological verification (HV%)**
- * **DCO 'death certificate only' (%DCO, %DCN)**

اعتبار داده ها

reabstracting and recoding *

- * موردیابی مجدد
- * چکیده سازی مجدد
- * نمونه گیری تصادفی / هدفمند
- * درصد توافق، اشکالات رایج و کدهای صحیح
- * میزان توافق بین چکیده کننده ها از لحاظ تشخیص های ثبت شده

* شناسایی و اصلاح خطاها

* اصلاح قوانین

* آموزش

اعتبار داده‌ها

* **تکرارگیری**

* تکرارگیری به روش درست و کامل متناسب با این راهنما

* **کنترل سازگاری داخلی internal consistency**

* سازگاری داده‌های مختلف با یکدیگر

* سازگاری بین تشخیص/سن، تاریخ بروز/ تاریخ تولد، بیماری؛ غیره

کامل بودن داده‌ها

درصد داده‌های مفقود برای عناصر اطلاعاتی:

* سن

* جنس

* آدرس

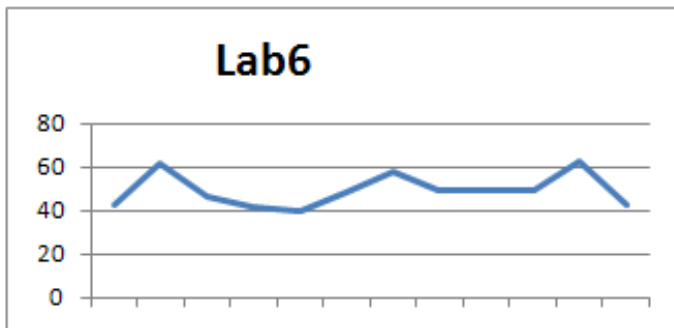
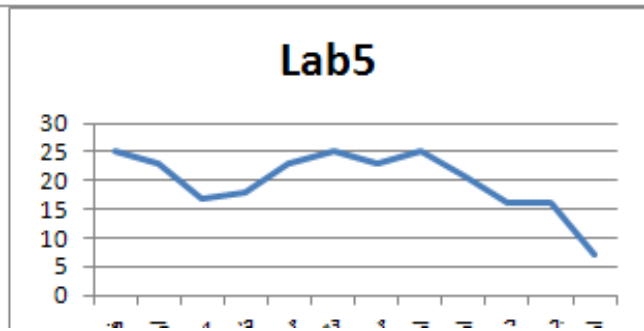
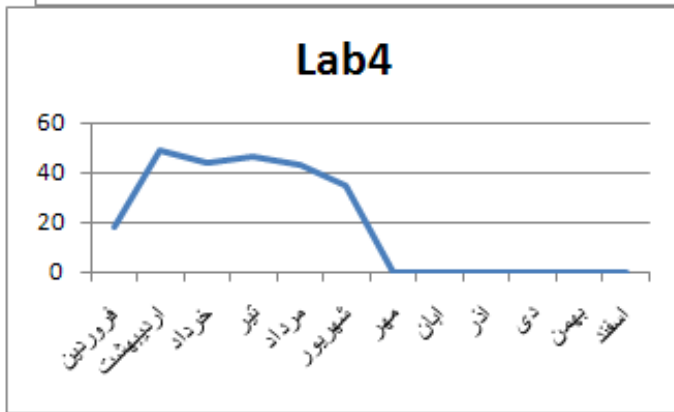
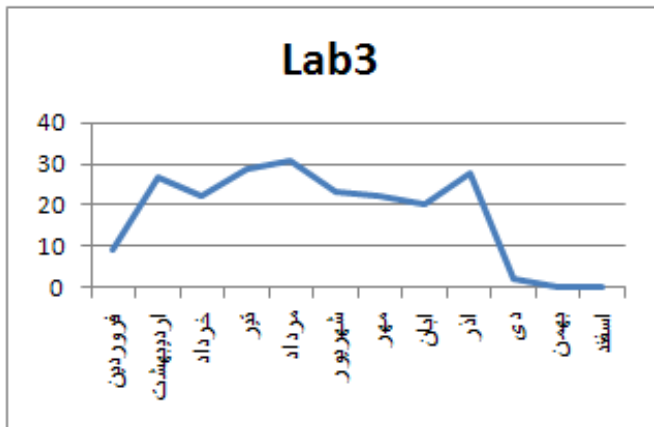
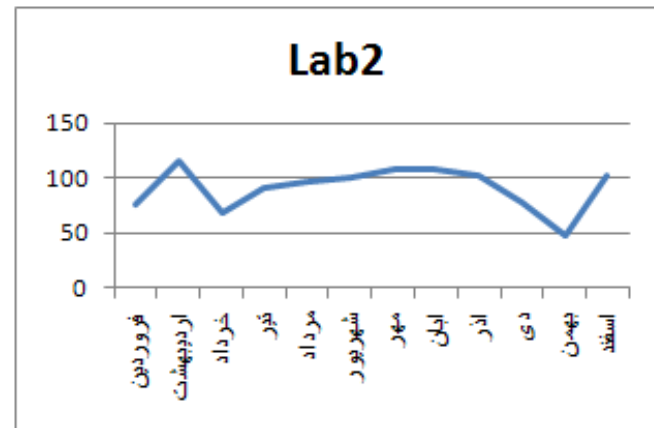
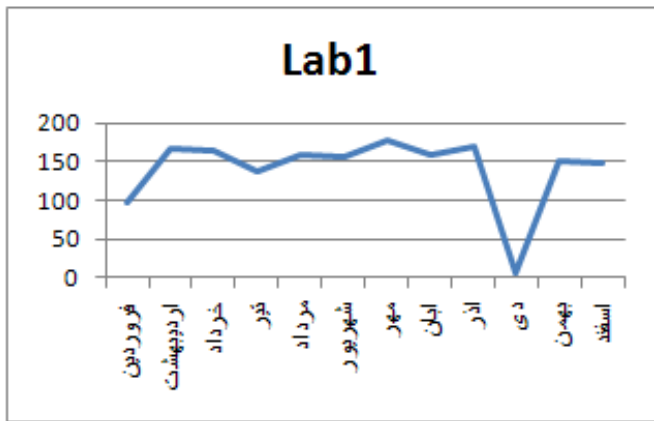
* درصد مواضع نامشخص

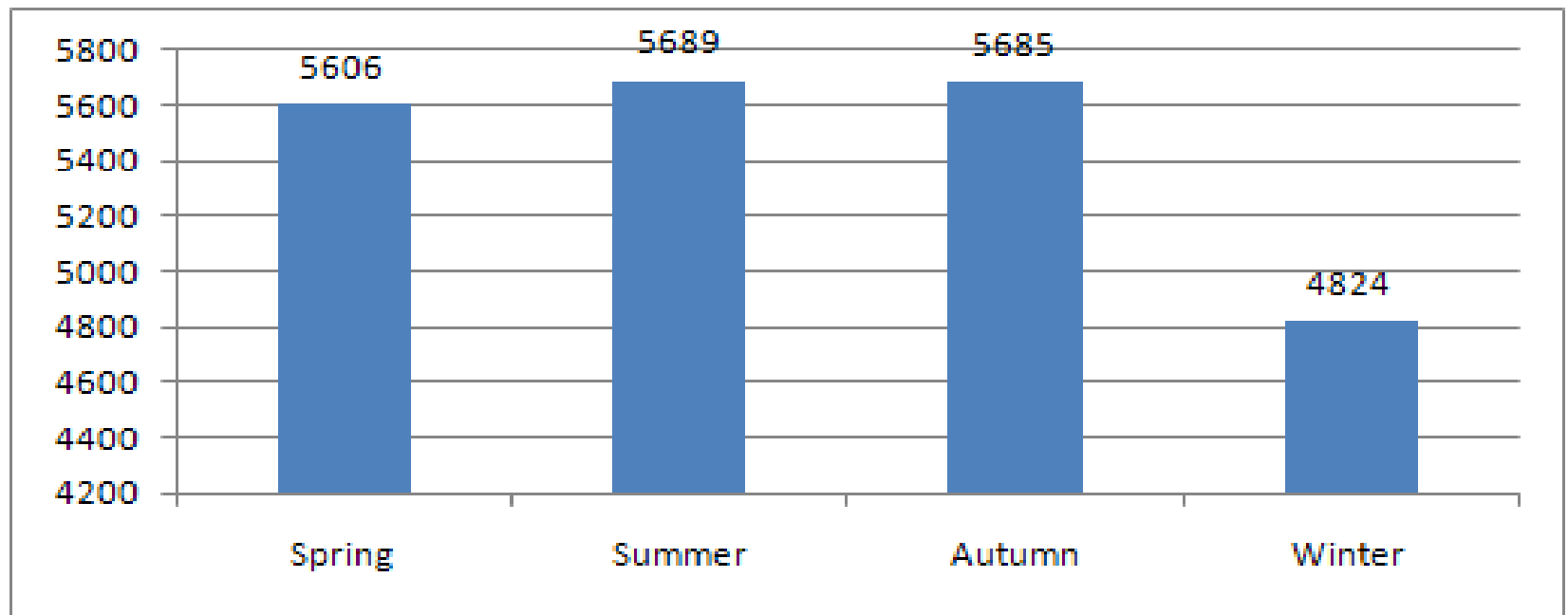
کامل بودن پوشش

Historic data methods *

- * ثبات در میزان ثبت در مقاطع زمانی stability
- * مقایسه میزان‌های بروز به دست آمده با میزان‌های مربوط به دوره‌های قبل
- * مقایسه بروز در جمعیت‌های مختلف (سن و جنس)

Figure 6: pattern of reporting of cancer cases to the National Cancer Registry by month in 6 laboratories in Tehran Provinces.





کامل بودن پوشش

histological verification (HV%) *

*** HV% too high = under-registration**

کامل بودن پوشش

independent case ascertainment *

active and passive registration *

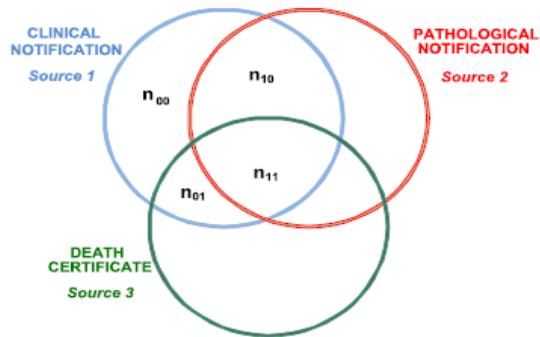
case- finding study *

number of sources/notifications per case *

* میانگین تعداد منبع اطلاعاتی به ازاء هر مورد شناسایی شده

کامل بودن پوشش

Capture–recapture methods *



		Source 3	
		Yes	No
Source 2	Yes	n_{11}	n_{10}
	No	n_{01}	n_{00}

Total = n_{++}

$$\hat{n}_{00} = \frac{n_{10} \cdot n_{01}}{n_{11}}$$

$$\hat{n}_{++} = (n_{10} + n_{11}) \cdot \frac{(n_{01} + n_{11})}{n_{11}}$$

$$Comp_{known} = \frac{n_{11} + n_{10} + n_{01}}{n_{++}}$$

$$Comp_{estimated} = \frac{n_{11} + n_{10} + n_{01}}{\hat{n}_{++}}$$

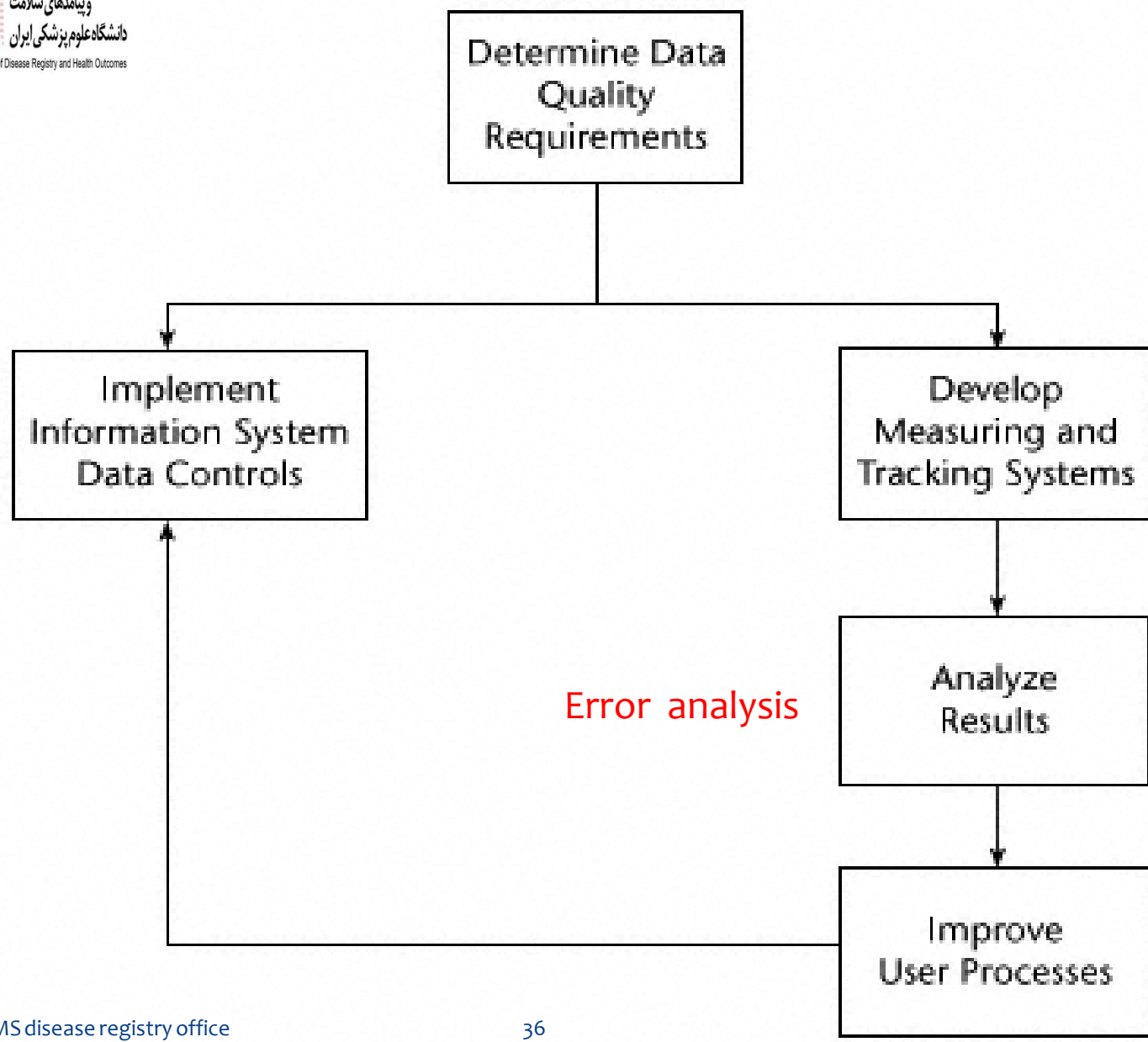
استانداردهای کیفیت داده

شاخص	
اعتبار و صحت داده‌ها	حداقل ۸۵-۹۵ درصد*
مواضع بد تعریف شده	کمتر از ۲۰ درصد
میزان کامل بودن پوشش	حداقل ۹۰ درصد*
میزان DCO	کمتر از ۲۰ درصد
موارد تکراری حذف نشده	کمتر از ۲ در هر هزار مورد*
داده‌های Missing برای سن	کمتر از ۲۰ درصد
منبع اطلاعاتی نامشخص	کمتر از ۲۰ درصد
داده‌های Missing برای جنس و محل سکونت (حداقل شهر)	کمتر از ۳ درصد*

بهنگام بودن

* فاصله زمانی بین اتمام سال رجیستری تا ارائه گزارش

* حداکثر تا ۲ سال (ثبت سرطان جمعیتی ملی)



Error types based on cause

1. Systematic errors

errors that can be attributed to a flaw or discrepancy in adherence to standard operating procedures or systems

- * E.g. Encoding software or improper training

2. Random Errors

- * Careless
- * Handwriting

TABLE 2.2. SOME CAUSES OF POOR HEALTH CARE DATA QUALITY.

Systematic	Random
Unclear data definitions	Illegible handwriting in data source
Unclear data collection guidelines	Typing errors
Poor interface design	Lack of motivation
Programming errors	Frequent personnel turnover
Incomplete data source	Calculation errors (not built into the system)
Unsuitable data format in the source	
Data dictionary is lacking or not available	
Data dictionary is not adhered to	
Guidelines or protocols are not adhered to	
Lack of sufficient data checks	
No system for correcting detected data errors	
No control over adherence to guidelines and data definitions	

Data Error Prevention

- Compose a minimum set of necessary data items
- Define data and data characteristics in a data dictionary
- Develop a data collection protocol
- Create user friendly data entry forms or interface
- Compose data checks
- Create a quality assurance plan
- Train and motivate users

Data Error Detection

- Perform automatic data checks
- Perform data quality audits
- Review data collection protocols and procedures
- Check inter- and intraobserver variability (if appropriate)
- Visually inspect completed forms (online or otherwise)
- Routinely check completeness of data entry

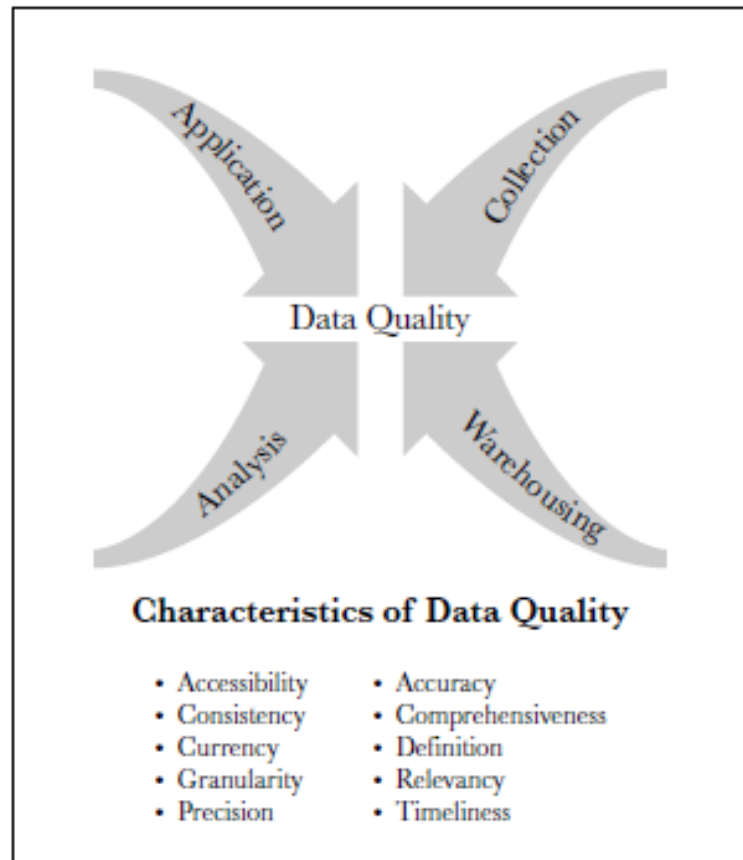
Actions for Data Quality Improvement

- Provide data quality reports to users
- Correct inaccurate data and fill in incomplete data detected
- Control user correction of data errors
- Give feedback of data quality results and recommendations
- Resolve identified causes of data errors
- Implement identified system changes
- Communicate with users

اقدامات

* پاک سازی داده به صورت روتین
* دستی / اتوماتیکی

- * missing values
- * incorrect
- * out-of-range values
- * inconsistent



Application – The purpose for which the data are collected.

Collection – The processes by which data elements are accumulated.

Warehousing – Processes and systems used to archive data and data journals.

Analysis – The process of translating data into information utilized for an application.



Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf



Development of a tool for comprehensive evaluation of population-based cancer registries



Abbas Sheikhtaheri^{a,b}, Azin Nahvijou^c, Zahra Sedighi^c, Maryam Hadji^c, Mohammad Golmahi^c, Gholamreza Roshandel^d, Omid Beiki^{e,f}, Zahra Ravankhah^g, Kazem Zendehtdel^{c,*}

^a Health Management and Economics Research Center, Iran University of Medical Sciences, Tehran, Islamic Republic of Iran

^b Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Islamic Republic of Iran

^c Cancer Research Center of Cancer Institute, Tehran University of Medical Sciences, Tehran, Islamic Republic of Iran

^d Golestan Research Center of Gastroenterology and Hepatology, Golestan University of Medical Sciences, Gorgan, Islamic Republic of Iran

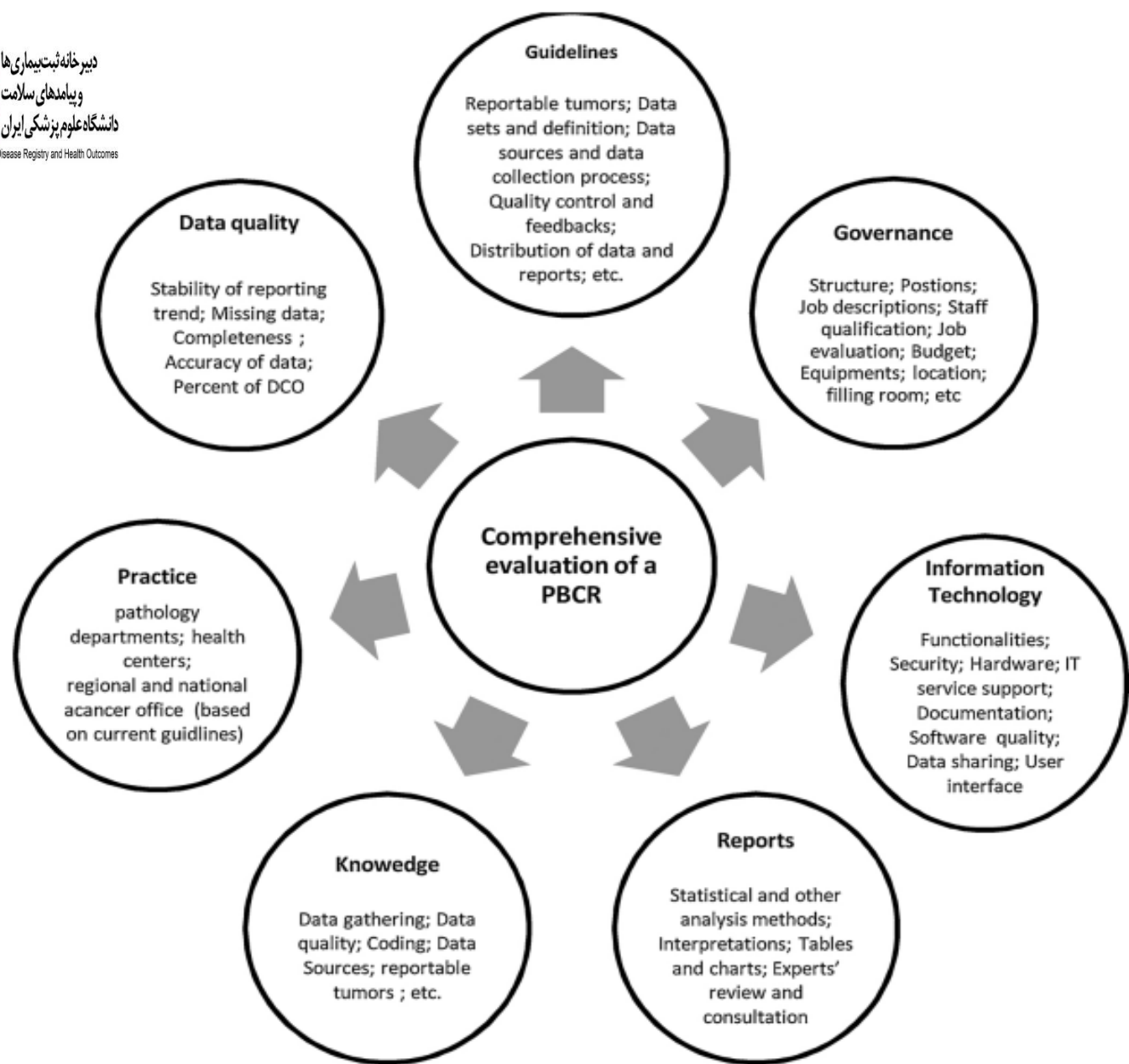
^e Kermanshah University of Medical Sciences, Kermanshah, Islamic Republic of Iran

^f Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

^g CDC, Isfahan Province Health Center, Isfahan University of Medical Sciences, Isfahan, Islamic Republic of Iran

ARTICLE INFO

ABSTRACT





Thank you for your attention

Any comments and questions?