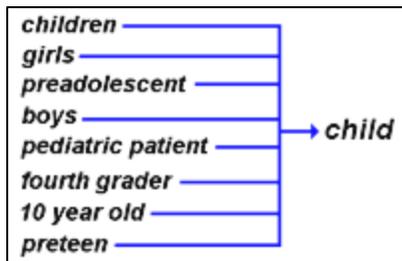# Subject Indexing and Controlled Vocabularies

Because "**natural**" languages cause so many headaches for information retrieval, **artificial indexing languages** or **controlled vocabularies** were created to address the problems created by natural language.

Subject indexing using controlled vocabularies is performed by people called *indexers*. Indexers review an item, and then assign enough indexing terms from the controlled vocabulary to describe the item. The indexing terms are added to a separate field in the record. The generic name for this field is the *descriptor field*. Ironically, there are quite a few other names for this field!

Example: Comparing the Indexer's 'description' using a controlled vocabulary with the author's description.

Author's description of his study[*]	Indexer's description of the paper using just the Medical Subject Headings.
Empathy goes a long way in weight loss discussions. This study explores how weight-related topics are discussed between physicians and their overweight and obese female patients.	• Empathy • Physician-Patient Relations • Counseling/methods • Obesity/psychology • Obesity/therapy • Adult • Female

Concepts in the author's description include the some of the following:
- *Empathy*
- *Discussions between physicians and their patients*
- *Female patients*
- *Weight loss / Weight related topics*

Can you match these to the heading assigned by the indexer?
Does it take more than one heading to match a concept mentioned by the author?
Are any of the headings more succinct than the author's description?

[*]Using the title and first sentence of the abstract

Controlled vocabularies provide four functions:

1. They **standardize vocabulary** by using a single word or phrase to represent a concept or subject
2. They **define topics** or subjects to reduce ambiguity
3. They **standardize phrasing**
4. They **pre-coordinate** (combine) **topics**.

Standardization of Vocabulary

In a controlled vocabulary system, one term or phrase is selected for a subject or concept and all the indexing of that topic must use that standard word or phrase.

Example 1.1:

The Medical Subject Headings (MeSH), the controlled vocabulary used to index Medline records, uses "**Pessure Ulcer**" as the standard term for that subject.

Although care is taken in selecting a heading so that it doesn't reflect obscure or archaic language, it is never-the-less an arbitrary selection. In the case of *Pressure Ulcer*, "Bedsore", "Decubitus Ulcer", or "Pressure Sore" could have been used instead, but they weren't.

The following excerpts from the Medline database show the title and MeSH fields of sixteen records indexed with the heading *Pressure Ulcer*. Note the variations in the authors' vocabulary *(title field)* including plural and singular word forms. Note that only two authors used the exact phrase *pressure ulcer* in their titles, although several used the plural form (*pressure ulcers*)

TI: **Pressure ulcer** prevention
MeSH: Databases, Factual*; Quality Indicators, Health Care*; Pressure Ulcer/*prevention & control; Humans; United States

TI: **Pressure sore** on malar prominences by horseshoe headrest in prone position
MeSH: Beds/*adverse effects; Pressure Ulcer/*etiology; Zygoma/*injuries; Anesthesia, General/methods; Atlanto-Axial Joint/surgery; Child; Female; Humans; Prone Position/physiology

TI: **Pressure ulcers** and their treatment and effects on quality of life: hospital inpatient perspectives
MeSH: Patient-Centered Care/*methods; Pressure Ulcer/*nursing; Adult; Aged; Aged, 80 and over; Female; Hospitalization; Humans; Male; Middle Aged; Patient Satisfaction; Patient-Centered Care/manpower; Patient-Centered Care/standards; Pressure Ulcer/drug therapy; Pressure Ulcer/psychology; Quality of Life/psychology

TI: Collaborative clinical quality improvement for **pressure ulcers** in nursing homes
MeSH: Nursing Homes/*statistics & numerical data; Pressure Ulcer/*epidemiology; Quality Assurance, Health Care/*trends; Humans; Pressure Ulcer/classification; Pressure Ulcer/prevention & control; Prevalence; Quality Assurance, Health Care/statistics & numerical data; Severity of Illness Index; United States/epidemiology

TI: The Norton, Waterlow, Braden, and Care Dependency Scales: comparing their validity when identifying patients' **pressure sore** risk
MeSH: Pressure Ulcer/*epidemiology; Humans; Nursing Assessment; Reproducibility of Results; Risk Assessment

TI: The prevention and treatment of **pressure ulcers**
MeSH: Pressure Ulcer/*therapy; Spinal Cord Injuries/*complications; Algorithms; Bed Rest; Humans; Hydrotherapy; Image Processing, Computer-Assisted; Nanotechnology; Pressure Ulcer/etiology; Pressure Ulcer/prevention & control; Telemedicine; Wheelchairs

TI: The perforator-sparing buttock rotation flap for coverage of **pressure sores**
MeSH: Pressure Ulcer/*surgery; Reconstructive Surgical Procedures/*methods; Surgical Flaps/*blood supply; Wound Healing/*physiology; Adult; Buttocks; Cohort Studies; Female;

Follow-Up Studies; Humans; Male; Middle Aged; Pressure Ulcer/diagnosis; Risk Assessment; Rotation; Sacrococcygeal Region; Severity of Illness Index; Treatment Outcome

TI: Nutrition also plays a role in **pressure ulcer** therapy
MeSH: Malnutrition/*prevention & control; Prealbumin/*physiology; Pressure Ulcer/*prevention & control; Wound Healing/*physiology; Humans; Malnutrition/complications; Malnutrition/diagnosis; Nutrition Assessment; Predictive Value of Tests; Pressure Ulcer/etiology

TI: Factors predicting cervical collar-related **decubitus ulceration** in major trauma patients
MeSH: Cervical Vertebrae/*pathology; Orthotic Devices/*adverse effects; Pressure Ulcer/*epidemiology; Pressure Ulcer/*etiology; Spinal Cord Injuries/*pathology; Adolescent; Adult; Aged; Aged, 80 and over; Female; Humans; Incidence; Male; Middle Aged; Predictive Value of Tests; Restraint, Physical/methods; Retrospective Studies; Risk Factors; Spinal Cord Injuries/therapy

TI: The causation and prevention of **bed sores**
MeSH: Pressure Ulcer/*etiology; Pressure Ulcer/*prevention & control; Humans; Necrosis; Posture; Pressure Ulcer/pathology; Risk Factors; **Skin**/blood supply; **Skin**/pathology

TI: **Decubitus ulcers**: a review of the literature
MeSH: Pressure Ulcer/*pathology; Pressure Ulcer/*therapy; Humans; Severity of Illness Index

TI: **Decubitus ulcers** after instituting epidural analgesia for pain relief in labour
MeSH: Analgesia, Epidural/*adverse effects; Analgesia, Obstetrical/*adverse effects; Pressure Ulcer/*etiology; Adjuvants, Anesthesia/administration & dosage; Anesthetics, Local/administration & dosage; Bupivacaine/administration & dosage; Female; Fentanyl/administration & dosage; Humans; Pregnancy; Pressure Ulcer/prevention & control

TI: **Skin sores** after spinal cord injury: relationship to life adjustment
MeSH: Adaptation, Psychological/*physiology; Pressure Ulcer/*etiology; Pressure Ulcer/*psychology; Spinal Cord Injuries/*complications; Spinal Cord Injuries/*psychology; Adult; Female; Humans; Male; Pressure Ulcer/epidemiology; Questionnaires; Treatment Outcome

TI: The treatment of **decubitus ulcers**: a century of misinformation in the textbooks
MeSH: Pressure Ulcer/*therapy; Anti-Infective Agents/therapeutic use; Clinical Competence; Diet; Humans; Massage; Nursing Evaluation Research; Physical Stimulation; Physical Therapy Modalities/methods; Pressure Ulcer/nursing; Pressure Ulcer/physiopathology; Wound Healing; Wound Infection/therapy

TI: Three nursing home patients who, due to inadequate care and poor nutrition, developed severe **bed sores**
MeSH: Nursing Homes*; Quality Assurance, Health Care*; Health Personnel/*education; Pressure Ulcer/*prevention & control; Humans; Periodicals

What's the point?
If you use the MeSH heading "*Pressure Ulcer*", you don't need to search all the other synonyms!

Example 1.2:

The standard term in Mesh for cancer is **neoplasms**. It is used very consistently, especially in headings for cancers by anatomical location:

Abdominal Neoplasms	Kidney Neoplasms
Breast Neoplasms	Muscle Neoplasms
Brain Neoplasms	Parotid Neoplasms
Eyelid Neoplasms	Skin Neoplasms

Cancers identified by histological (cell/tissue) type have distinct names:

Adenocarcinoma	Lymphoma
Hepatoblastoma	Osteosarcoma
Leukemia	Sarcoma

A paper about a cancer of a histological type at a specific anatomical location will be indexed using both types of cancer headings:

TI: Transhiatal esophagectomy for **squamous cell carcinoma of the esophagus**.
MeSH: **Carcinoma, Squamous Cell**/*surgery; **Esophageal Neoplasms**/*surgery; Esophagectomy/*methods; Adult; Aged; **Carcinoma, Squamous Cell**/mortality; **Esophageal Neoplasms**/mortality; Esophagectomy/adverse effects; Esophagectomy/mortality; Female; Humans; Male; Middle Aged; Palliative Care; Postoperative Complications; Survival Rate

TI: Trends in **oesophageal cancer** incidence and mortality in Europe.
MeSH: Adenocarcinoma/*epidemiology; **Carcinoma, Squamous Cell**/*epidemiology; **Esophageal Neoplasms**/*epidemiology; Mortality/*trends; Adult; Alcohol Drinking/adverse effects; Europe/epidemiology; Female; Humans; Incidence; Male; Middle Aged; Smoking/adverse effects; Survival Rate

Defining the Scope of Topics

When there might be ambiguity about the meaning of a term, controlled vocabularies will often define how the term is to be used. Look at these two examples from MeSH:

Example 2.1:
(excerpt from MeSH)

> **Child** *(MeSH heading)*
> age 6-12 years
> **Child, Preschool** *(MeSH heading)*
> age 2-5 years

Example 2.2:
(excerpt from MeSH)

> **Language Development Disorders** *(MeSH heading)*
> do not confuse with Language Disorders, disorders of use or comprehension of language: **Language Development Disorders is "failure to understand or speak the language at the expected age"**.

Standardized Phrasing

Controlled vocabularies may use multiword descriptors:

- Muscle Spasticity
- Clinical Laboratory Information Systems
- Tuberculosis, Hepatic

Because controlled vocabularies are often used in print indexes (at least historically) as well as computerized databases the word order is often inverted so that related headings are grouped alphabetically:

- Pancreatitis
- Pancreatitis, Acute Necrotizing
- Pancreatitis, Alcoholic

In a printed index, the inverted phrase brings all the *pancreatitis* papers together in one place. If the normal word order was maintained the papers on various kinds of pancreatitis would be split by perhaps hundreds of pages or even by multiple volumes in a large annual index. (Think of how these would be distributed in a multivolume encyclopedia.)


Pre-coordination of Subjects

Usually controlled vocabulary terms represent a single concept, but they may be multi-faceted, representing more than one concept:

MeSH Heading	Venn Diagram
"Housing for the Elderly" *Used instead of assigning 2 headings: "Housing" and "Aged."*	Elderly — Housing
"Lung Transplantation" *Used instead of assigning the heading "Lung" qualified with the subheading "transplantation."*	Lung — Transplantation
"Injections, Spinal" *Used instead of assigning 2 headings: "Injections" and "Spine".*	Spine — Injections

Some indexing systems add qualifiers or subheadings to the main indexing term. These qualifiers help the indexer precisely define the topic. The MeSH indexing system includes nearly 90 topical subheadings. Three examples of MeSH *heading* **-** *subheading* combinations are shown below:

- Saliva / **chemistry**
- Lung Diseases,Obstructive / **nursing**
- Sufentanil / **pharmacology**

Although these multi-faceted headings could be represented using an "AND" Venn diagram, it might be more helpful to think of the main-heading/subheading relationship as a subset. In the third example, "Sufentanil - **pharmacology**", the indexer has indicated that it is the pharmacology of sulfentanil that is discussed rather than its adverse effects, therapeutic use, etc:
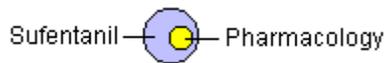
The following excerpts from the Medline database show the title and MeSH fields of three records indexed with the heading *Sufentanil*. Read each title and look at the MeSH headings. Note how the addition of the subheading helps to define how Sufentanil is discussed in each paper:

TI: Low-dose sufentanil increases cerebrospinal fluid pressure in human volunteers.
MeSH: Cerebrospinal Fluid Pressure / drug effects
     Sufentanil / pharmacology

TI: Respiratory depression after 5 micrograms of intrathecal sufentanil.
MeSH: Analgesics, Opioid / adverse effects
     Respiratory Insufficiency / chemically induced
     Sufentanil / adverse effects

TI: Simultaneous extraction of sufentanil and midazolam from human plasma.
MeSH: Midazolam / blood
     Midazolam / isolation and purification
     Sufentanil / blood
     Sufentanil / isolation and purification