**Building a Comprehensive Breast Implant Ontology Leveraging GUDID and Unstructured Data Sources**

**Abstract**

**Background:**  Individuals with breast implants have a risk of developing breast implant-associated anaplastic large cell lymphoma (BIA-ALCL), which based on published literature, appears to be associated with breast implants that have a textured surface. The FDA product codes and the Global Medical Device Nomenclature (GMDN) do not capture breast implants’ surface type. This significantly limits our ability to assess the potential association between the breast implant surface types and BIA-ALCL. A Breast Implant Ontology can categorize all the breast implant products and their attributes under an ontological structure, which can be used by a semantic reasoner tool to automatically classify breast implants features (e.g. smooth or textured surface types). This ontology could also be used as a backend dictionary for a text mining tool to help explore the patterns/trends from BIA-ALCL cases reported through the Medical Device Reporting System or other data sources.

**Objective:** The objective of this project is to create an extensive ontology to categorize the different breast implants and their product features, including but not limited to: manufacturer, brand, filling, and so on. The ontology will help FDA determine possible relationships between BI structures and patient symptoms that correspond with BIA-ALCL.

**Methods:** A theoretical design of the Breast Implant Ontology was developed prior to ontology development. A total of 1,739 breast implant data, consisting of 33 unique fields including – but not limited to –device identification (DI), record status, publish date, brand name, model number, company name, and device description was downloaded from GUDID database on June 28th, 2018. Additional information was also retrieved, including a unique device name, device dimensions (width, height, projection, etc) from sponsor’s catalogs, and other PMA approval orders. Overarching classification categories that were eventually included are device manufacturer, brand, style, filling, profile/dimensions, size, shape, shell, shell surface, and product code. Data such as DI, model number, catalog number, manufacturer’s device description, GMDN name, and GMDN definition are also included as annotations for each individual device.

**Results:** The ontology is searchable with SPARQL queries or Protégé’s DL Query tab, allowing users to find individual data as well as their relationships. This may simplify text mining in the context of discovering possible correlations between certain types of breast implants and BIA-ALCL. The ontology is also available as open source to aid other researchers and organizations.

**Keywords**

Breast implants, Ontology, Automatic classification, ALCL, OWL

**Background**

The Background section should explain the background to the study, its aims, a summary of the existing literature and why this study was necessary.

**Methods**

The methods section should include:

* the aim, design and setting of the study
  + The aim of this project was to compile all available breast implant data into a domain ontology that would be primarily used as a data analysis tool.
  + All materials were electronic. As such, the ontology building was done entirely on computers. FDA provided Dell Latitude 5580 with a CORE i7 processor and 16.0 GB RAM.
* the characteristics of participants or description of materials
  + Global Unique Device Identification Database (GUDID) – GUDID is an [FDA managed database](https://www.fda.gov/medicaldevices/deviceregulationandguidance/uniquedeviceidentification/globaludidatabasegudid/default.htm) that stores information on every medical device with and identifier.
    - Using the query: “productCodes.fdaProductCode.productCode:(FWM) OR productCodes.fdaProductCode.productCode:(FTR)”, we found all saline-filled (FWM) and silicone gel-filled (FTR) breast implant data
    - We exported all results, giving a folder with delimited text files, with a total column count of 47
    - They are as follows: PrimaryDI, deviceRecordStatus, devicePublishDate, deviceCommDistributionEndDate, deviceCommDistributionStatus, brandName, versionModelNumber, catalogNumber, companyName, deviceCount, deviceDescription, DMExempt, premarketExempt, deviceHCTP, deviceKit, deviceCombinationProduct, singleUse, lotBatch, serialNumber, manufacturingDate, expirationDate, donationIdNumber, labledContainsNRL, labledNoNRL, MRISafetyStatus, rx otc, deviceSterile, sterilizationPriorToUse, sizeType, size (Unit), size (Value), sizeText, productCode, productCodeName, deviceId, deviceIdType, deviceIdIssuingAgency, containsDINumber, pkgQuantity, pkgDiscontinueDate, pkgStatus, gmdnPTName, gmdnPTDefinition, phone, phoneExtension, email
    - In the process of analyzing the data, we discovered 6 devices listed as breast implants but were in fact saline pumps
    - The columns which we ended up taking directly from GUDID were PrimaryDI, devicePublishDate, brandName, versionModelNumber, catalogNumber, companyName, deviceDescription, sizeText, productCode, productCdoeName, deviceId, gmdnPTName, gmdnPTDefinition
    - Columns were left out primary because they were empty.
  + Company Catalogs – (Link to Each)
    - Sientra
    - Natrelle
    - Ideal Implant
    - Mentor
  + [FDA PMA Data](https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMA/pma.cfm) – “Premarket approval (PMA) is the FDA process of scientific and regulatory review to evaluate the safety and effectiveness of Class III medical devices. Class III devices are those that support or sustain human life, are of substantial importance in preventing impairment of human health, or which present a potential, unreasonable risk of illness or injury.” We downloaded all FDA PMA data concerning the 8 breast implants
* a clear description of all processes, interventions and comparisons. Generic names should generally be used. When proprietary brands are used in research, include the brand names in parentheses
* the type of statistical analysis used, including a power calculation if appropriate

**Results**

This should include the findings of the study including, if appropriate, results of statistical analysis which must be included either in the text or as tables and figures.

**Discussion**

For research articles this section should discuss the implications of the findings in context of existing research and highlight limitations of the study. For study protocols and methodology manuscripts this section should include a discussion of any practical or operational issues involved in performing the study and any issues not covered in other sections.

**Conclusions**

This should state clearly the main conclusions and provide an explanation of the importance and relevance of the study to the field.

**List of abbreviations**

If abbreviations are used in the text they should be defined in the text at first use, and a list of abbreviations can be provided.