Generating Hypotheses by Discovering Implicit Associations in the Literature: A Case Report of a Search for New Potential Therapeutic Uses for Thalidomide

By Marc Weeber, Henny Klein, Lolkje T. W. deJong – van den Berg, Alan Aronson, Grietje Molema

From J Am Med Inform Assoc, 2003

Presented by Nancy Baker

INLS 279

March 22, 2005

DAD Overview

- Automation on Swanson's A-B-C paradigm
- Uses PubMed citations – title and abstract (not just titles)
- Role of user vital in initial question, filtering
- User decides interestingness
- Strengthen or reject the hypothesis by looking at mechanisms/pathways

Key Innovation

- Units of analysis are UMLS metathesaurus concepts, not terms
- Why?
  - Only interested in biomedical concepts
  - Won't need stop words and words not relevant to medicine
  - Want to indentify and include compound terms (e.g. Blood Pressure)
  - UMLS concepts have semantic types – for abstraction and filtering
  - Multiple words collapse to one concept

UMLS examples

- IL-12, IL12, interleukin 12, CLMF, cytotoxic lymphocyte maturatin factor, natural killer cell stimulatory factor all refer to Interleukin-12
- Concepts have 134 categories: Disease or Syndrome, Gene or Genome, Amino Acid, Peptide, or Protein

Next steps (from earlier paper)

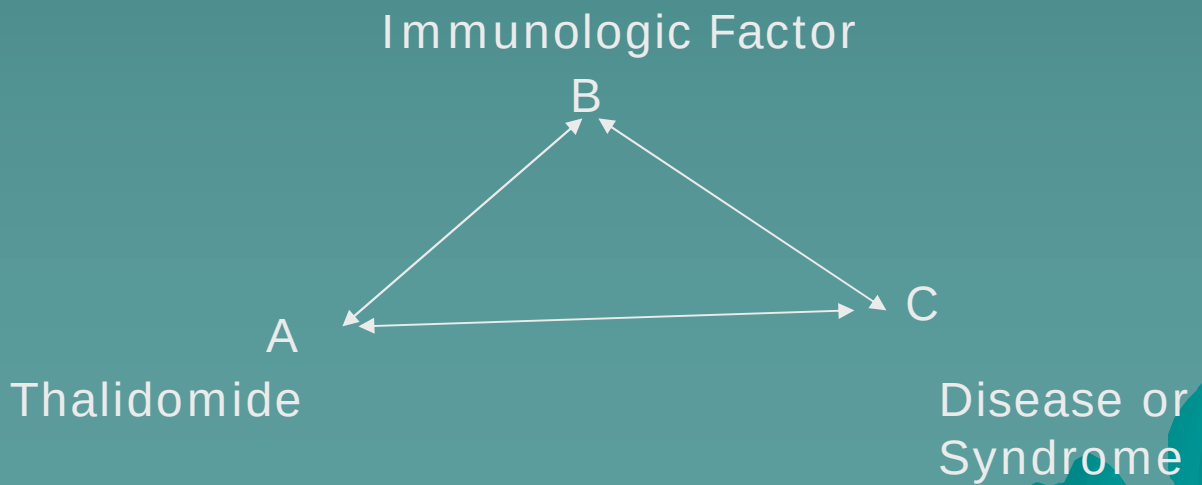
- Swanson's first discovery has been successfully simulated, so ...
- Adverse drug reactions
  - ADRs may benefit other conditions
    - Lots of examples (finasteride-alopecia)
    - DAD : drug-ADR-disease or disease-ADR-drug
    - Investigating retrospectively case of finasteride

Future Perspectives (from earlier paper)

- For some applications other routes are better than PubMed
  - Genetic databases, for instance
  - Combine text and database information

Case Report - Thalidomide

- Why thalidomide?
  - Known to have immunomodulatory and antiinflammatory properties
  - Anti-wasting (HIV)



The ABC discovery model.

The Discovery Process

- Experimental Setting: The people
  - Information Scientist (Weeber)
  - Pharmacologist/immunologist
  - Worked in collaboration

Generating Hypotheses

- Start with thalidomide
  - Pubmed search (titles and abstracts)using terms thalidomide, sedoval, synovir,kevadon
  - Downloaded results
  - Mapped results to UMLS concepts
  - Applied semantic filter
    - Selected only concepts classified as Immunologic Factor from sentences which also mentioned thalidomide.
    - “The increase of Interleukin-2 levels after application of thalidomide …”

Generating Hypotheses (cont.)

- In sentences with thalidomide, 3,860 concepts occurred, 82 with symantic type “Immunologic factor”
- Removed concepts considered too general
- Tool allows viewing of A-B
- Promising B concepts selected
  - Frequency
  - Expert knowledge

Results : immunologic factors

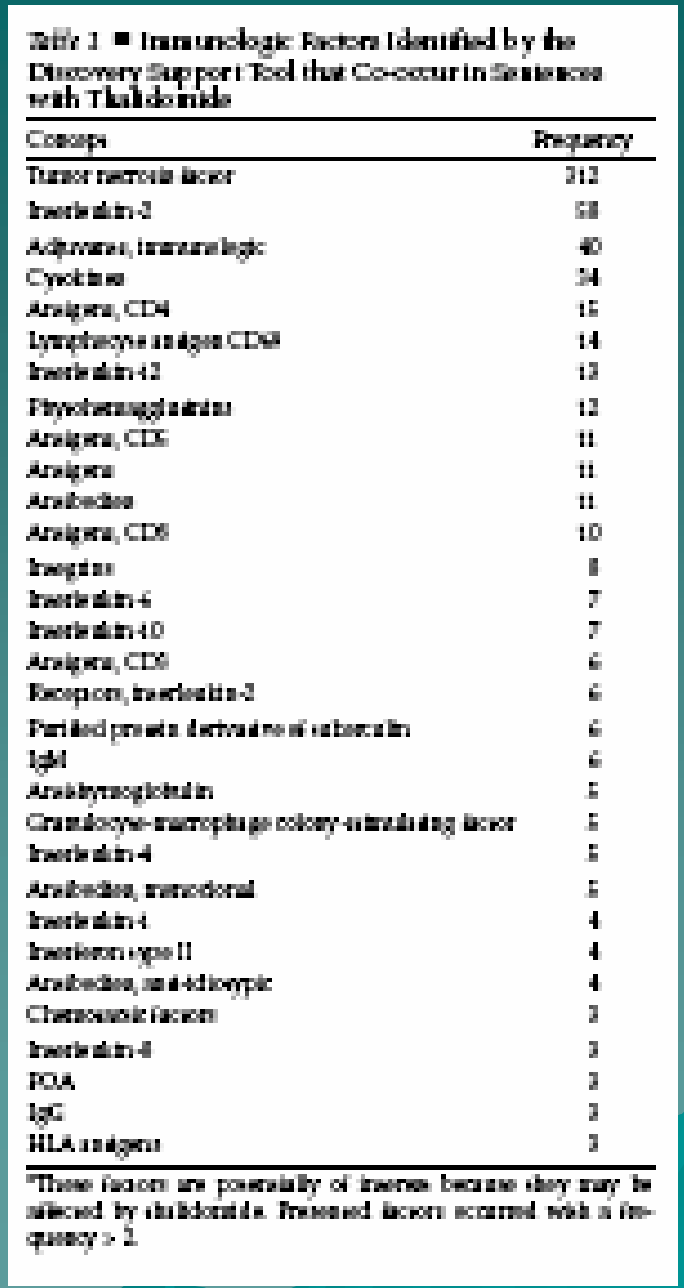


Table 1 ■ Immunologic Factors Identified by the Discovery Support Tool that Co-occur in Sentences with Thalidomide

Concept	Frequency
Tumor necrosis factor	312
Interleukin-2	58
Adjuvants, immunologic	40
Cytokines	34
Antigens, CD4	15
Lymphocyte antigen CD40	14
Interleukin-12	13
Phytohemagglutinins	12
Antigens, CD5	11
Antigens	11
Antibodies	11
Antigens, CD8	10
Integrins	8
Interleukin-6	7
Interleukin-10	7
Antigens, CD3	6
Receptors, Interleukin-2	6
Purified protein derivative of tuberculin	6
IgM	6
Antilymphocytic serum	5
Granulocyte-macrophage colony-stimulating factor	5
Interleukin-4	5
Antibodies, monoclonal	5
Interleukin-1	4
Interferon type II	4
Antibodies, anti-idiotypic	4
Chemotactic factors	3
Interleukin-8	3
FOA	3
IgG	3
HLA antigens	3

*These factors are potentially of interest because they may be affected by thalidomide. Presented factors occurred with a frequency > 2.

Results

- Domain expert selected Interleukin-12 and Interleukin-10.
- Thalidomide inhibits IL-12 and stimulates IL-10.
- Further research focuses on IL-12.

Generating Hypotheses (cont.)

- The selected B concepts were used as PubMed search criteria
- Diseases selected using semantic filtering

Interleukin-12

- 3,846 MEDLINE citations had concept Interleukin-12
- 420 Disease or Syndrome concepts co-occurred with interleukin-12
- Filtered
  - Threw out too general, too few occurrances, already known connections
  - Subjective

List of diseases

Table 2 ■ Twelve Diseases for Which Thalidomide May Be Used Therapeutically

Disease	No. of MEDLINE Citations
Acute pancreatitis	8,446
Atherosclerosis	15,873
Brucellosis	4,350
Chickenpox	6,304
Chronic hepatitis C	1,820
Crohn' disease	6,260
H. pylori-induced gastritis	7,108
Myasthenia gravis	6,228
Pulmonary fibrosis	1,602
Purpura	7,680
Sarcoidosis	706
Sjögren's syndrome	6,262

*These uses are not yet reported in the biomedical literature (PubMed) and are considered hypotheses that warrant further bibliographic investigation. The second column provides the number of citations to be analyzed.

Evaluating Hypotheses

- Download and analyze citations
- Looked at A-B concepts juxtaposed with B-C

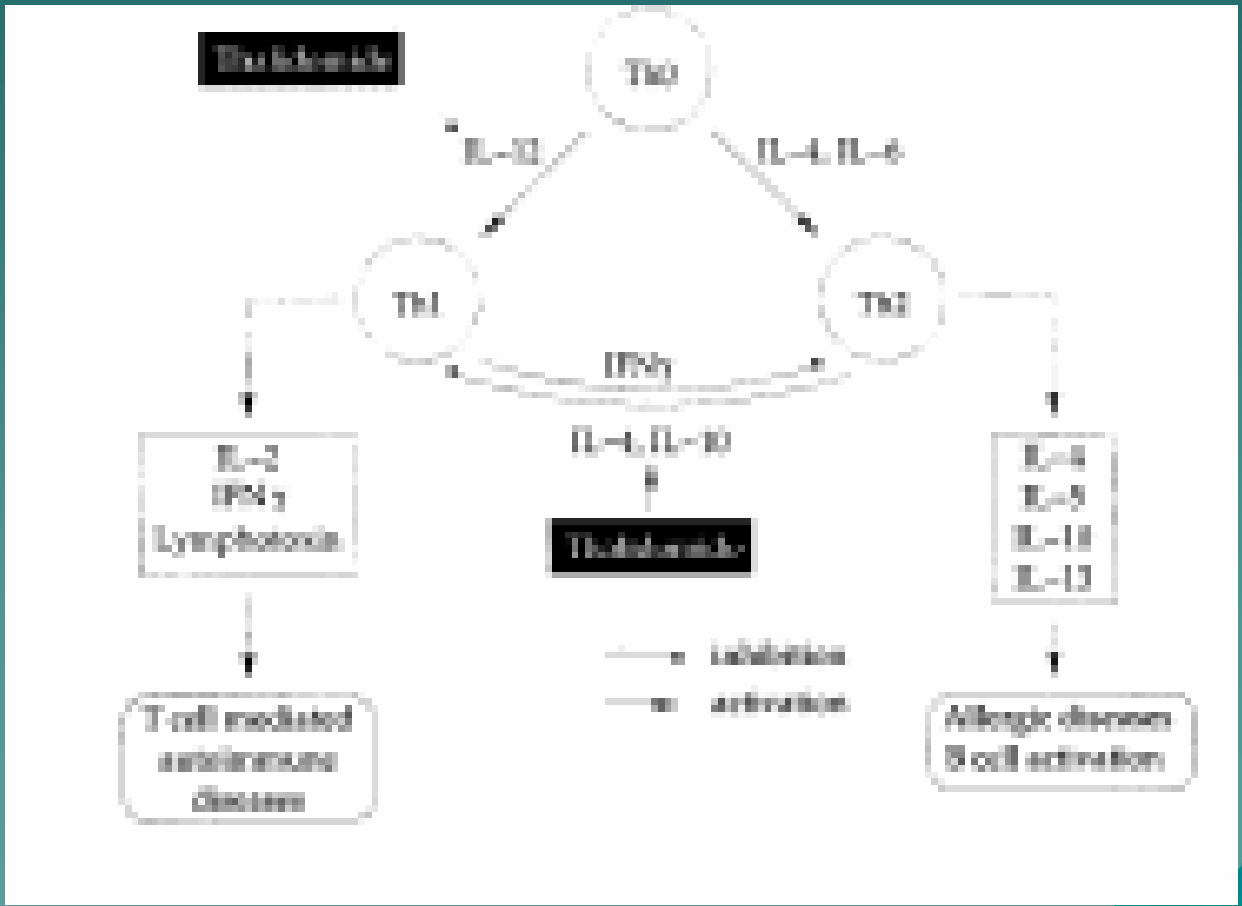
Results:

- Chronic Hepatitis C
  - Inflammatory disease of the liver
  - Th1/Th2 cytokine balance involved
- Myasthenia Gravis
  - Organ specific autoimmune disease affecting neuromuscular junctions
  - Aberrant production of cytokines

Results:

- *H. pylori* induced gastritis
  - Th1 medicated chronic inflammation
- Acute Pancreatis
  - Again, Th1 mediated response.

Thalidomide Th1/Th2



Other databases

- Queries other databases
  - Biological Abstracts
  - CINAHL – Nursing and Allied Health
  - EMBASE
  - Current Contents
  - Altavista and Google
- Some discussion of thalidomide and diseases – nothing definitive

Conclusion

- These four diseases represent novel potential targets for thalidomide
- Clinical investigation needed
- Although the computer system is valuable, discovery is an intellectually intensive process