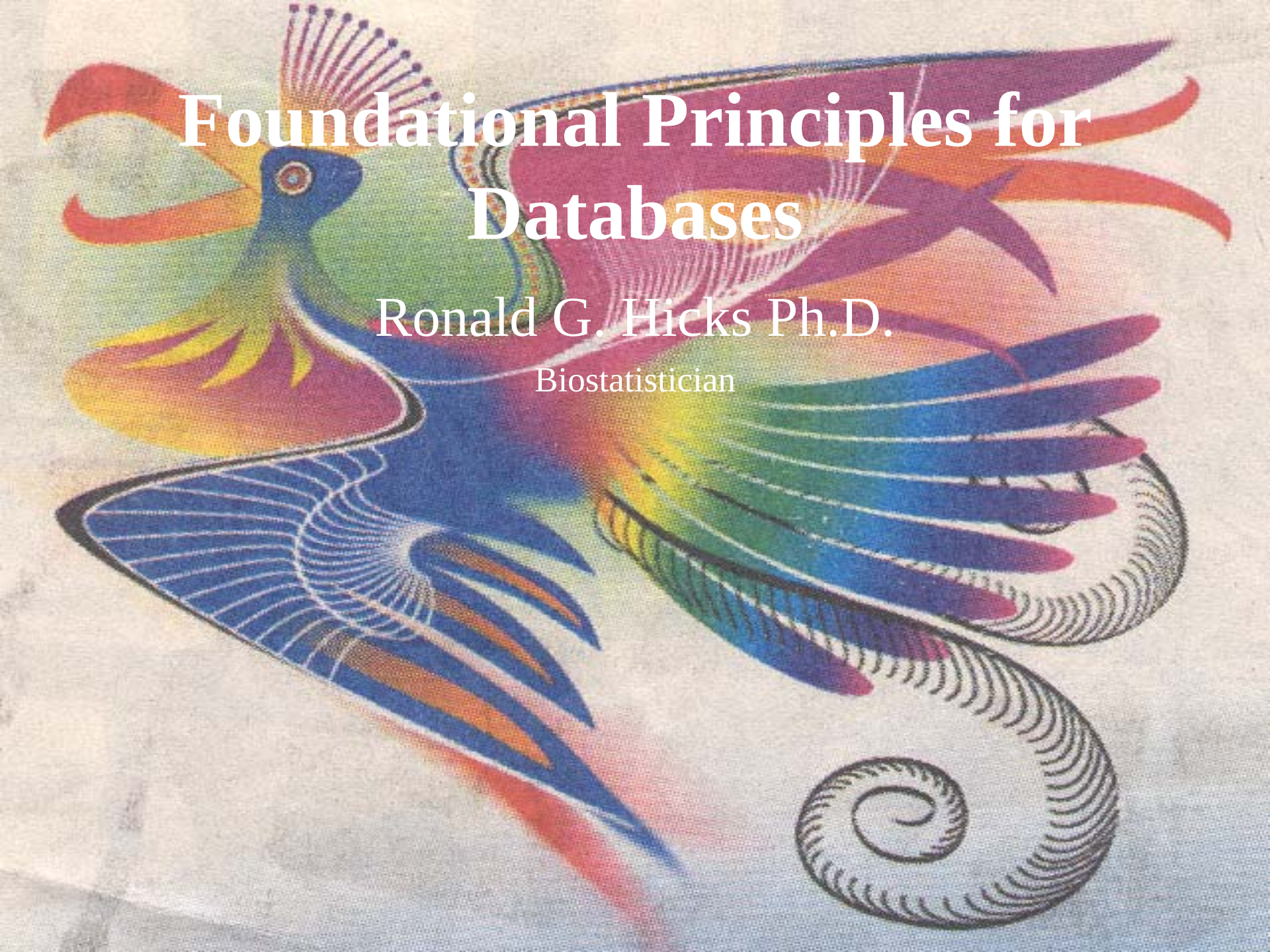

Foundational Principles for Databases

Ronald G. Hicks Ph.D.

Biostatistician

Database Semantics has Many Facets.

1. Dictionary Definitions = What is measured
2. Operational Definitions = How measured
   - Hypothetical Constructs
   - Expectancies
3. Internal Reliability
4. Scale
5. Inferences

Dictionary Definitions inform what data is entered into a particular column (variable).

For example Blood Pressure:

- Systolic measures occupies a column.
- Diastolic measures occupy another column.

On our way, nothing can be more direct and simple.

NOT!

Operational Definitions

Operational Definitions inform how the variable is to be measured.

BP taken on the right or left arm, if both, separate columns for both?

American or British technique?

Sitting for ten minute or just ran up a flight of stairs?

Automatic BP cuff?

Hypothetical Constructs

Hypothetical Constructs are concepts that can not be measured directly. For example:

Prior to 1953 the “Gene” was a hypothetical construct.

Other hypothetical constructs still exist; anxiety, depression, numerous categories of dementia.

Hypothetical Constructs

Hypothetical constructs needs additional semantic content in defining what and how the construct is assessed.

Expectancies

Assessors, taking the measurements, are not neutral. Assessors have internal expectations as to the results. “This patient has high BP” Or, “This patient is probably in the control group.

Expectancies effect the data being entered into the database, referred as internal reliability.

Reliability

Reliability is a measure to determine how consistent the data has been entered. How repeatable are the data-entries.

For example, did a second person independently measure BP.

A correlation of the two separate data-entries, r^2, yields an internal reliability coefficient.

Scale

All numbers are not created equal.

Some numbers are used to denote independent entities.

Diabetic, Yes = 1, No = 2. Nominal Scale, or,

Is HbA1c used as the data, interval scale.

Scale

Nominal scale denote different categories.

Ordinal scale denote categories with some form of direction, e.g., >, <.

Interval scale has direction of consistent size.

Ratio is internal scale with a true zero.

Inference

Inference is a concluding generalisation as to why the results have occurred.

The results, “So what conclusion” is generally the inference.

In general the tighter the semantics the more robust the resulting inference.

Semantic Inference

The lack of semantics, probably has lead to studies conflicting with each other.

If database interoperability is to be achieve semantic ontology must be achieved.

Primary prerequisite for the establishment of an Electronic Medical/Health Record.

What Can Go Wrong, Will Go Wrong

What happens went a database does not comply with the above suggestions.

An example of what can go wrong, a nom de guerre of “LOOKGOOD.”

LOOKGOOD did everything wrong.

LOOKGOOD was established late in the last century.

- LOOKGOOD wrote a simple dictionary of column terms in early 2006, but no operational definitions.
- LOOKGOOD has no data reliability confirmations.
- LOOKGOOD’s data columns have mixed scales.
- Hypothetical Constructs are not acknowledged.

LOOKGOOD

- LOOKGOOD did not acknowledge assessors expectancies effects.

- LOOKGOOD’s inferences?

LOOKGOOD'S Inferences

- LOOKGOOD created over two hundred administrative KPIs from data intended for client benefits.
- LOOKGOOD'S KPIs were charted and colour graphed automatically.
- LOOKGOOD'S KPIs were the archetype of "look good."

Statistical Analysis of LOOKGOOD

- The statistical analysis of LOOKGOOD results in a significance level of 0.99.

- The results are totally random.

LOOKGOOD'S TOUT

- LOOKGOOD'S "Data" was charted in graphic colour.
- LOOKGOOD'S "Data" in the form of KPIs was automatically update.
- LOOKGOOD'S "KPIs" were hyped and touted.

LOOKGOOD'S "Data"

- LOOKGOOD'S "Data" based KPIs was well dressed with:
- Eye shadow!
- Mascara!
- Ear rings!
- Hair styled!
- Tutu
- And lip-stick!

LOOKGOOD'S Resultant

- Removal of the eye shadow, and
- The Mascara,
- The Ear rings,
- The Hair styling,
- The Tutu, and
- The Lip-stick,
- The essences is found.

LOOKGOOD's Consequences

LOOKGOOD has consequences:

- The original purpose of LOOKGOOD has been changed.
- The LOOKGOOD hype has become institutionalised, a prime agenda.

Principles

- The point of this comic demonstration is to dramatise some the consequences of violating some of the basic principles of database construction.
- Institutions and people become vested in “their creation.” Band-Aids, hype, and retrospective changes are applied.
- Start with the basic principles and rest will not be necessary.

Simple Excel construction:

- Each Variable has a different column.
- Each Column has a name.
- Each Name is hyperlinked to its definition.

Var_Name	2VarName	**3VarName**	**4VarName**
Entry 1 ID	Data B1	Data C1	Data D1
Entry 2	Data B2	Data C3	Data D4
Entry 3	Data B3	Data C4	Data D5
"	"	"	"

Transfer

When the spreadsheet is transferred to a statistical package, accurate labels can be attached to the analysis's output.

	Name	Type	Width	Decimal	Label	Values	Missing	Columns	Align	Measure
1	etry1_id	Numeric	8	2	Variable 1 DSM10	None	None	8	Right	Nominal
2	var_name	Numeric	8	2	Variable 2 day of year	None	None	8	Right	Scale
3	varname2	Numeric	8	2	Variable 3 month of year	None	None	8	Right	Scale
4	varname3	Numeric	8	2	Variable 4 year	None	None	8	Right	Scale
5	varname4	Numeric	8	2	Variable 5 Systolic BP	None	None	8	Right	Scale

浴火鳳凰