# Orphans in Actuarial Data

DM-1: Data Cleansing for Predictive Modeling

Jeremy Benson

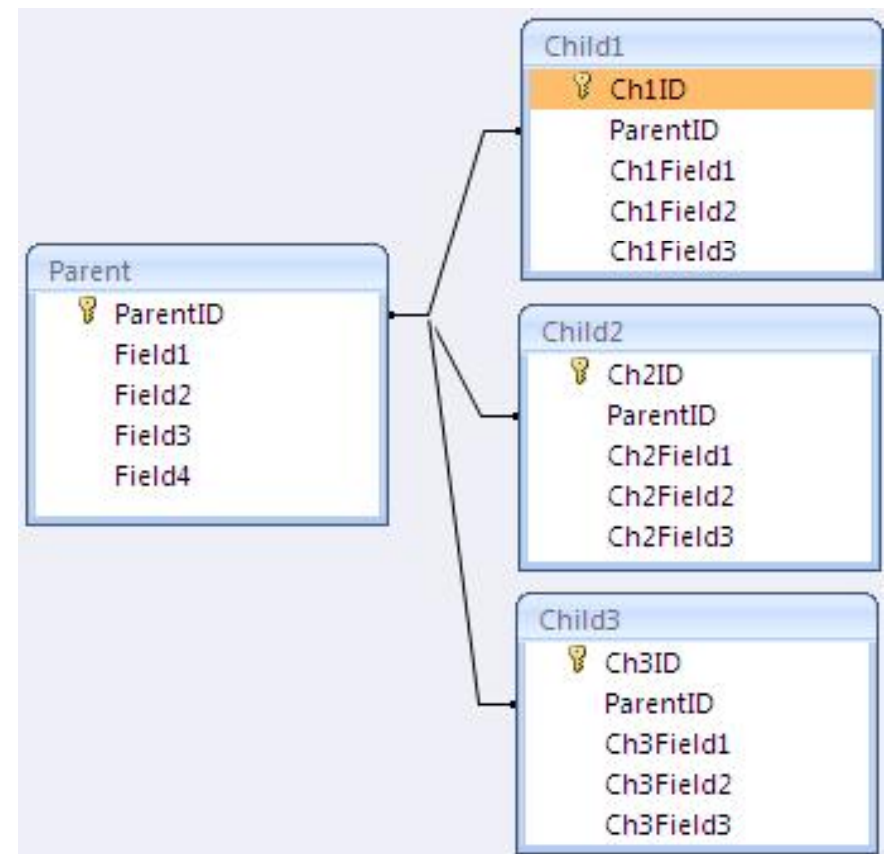March 2012 – RPM Seminar

# Antitrust Notice

- **The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.**

- **Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding − expressed or implied − that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.**

- **It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.**

# What is an Orphan?

- **Definition** − A record in a child table without an associated record in a parent table.

- **Parent and Child Tables**

| ParentID | Field1 | Field2 | Field3 | Field4 |
|----------|--------|--------|--------|--------|
| 1 | A | X | 1 | 5 |
| 3 | B | Y | 3 | 6 |
| 4 | C | Z | 5 | 7 |
| 5 | D | T | 7 | 8 |

| Ch1ID | ParentID | Ch1Field1 | Ch1Field2 | Ch1Field3 |
|-------|----------|-----------|-----------|-----------|
| 1 | 1 | R | 1 | N |
| 2 | 1 | S | 2 | N |
| 3 | 2 | R | 1 | Y |
| 4 | 2 | S | 4 | N |
| 5 | 3 | R | 4 | Y |

Child1
- Ch1ID
- ParentID
- Ch1Field1
- Ch1Field2
- Ch1Field3

Parent
- ParentID
- Field1
- Field2
- Field3
- Field4

Child2
- Ch2ID
- ParentID
- Ch2Field1
- Ch2Field2
- Ch2Field3

Child3
- Ch3ID
- ParentID
- Ch3Field1
- Ch3Field2
- Ch3Field3

## Orphans in Actuarial Data

■ **What orphans are there in actuarial data?**

- − A claim with no associated policy

- − Premium with no associated policy

- − A code with no value in the code lookup table

- − A customer not in the master customer table

- − An agent not in the master agent table

- − A claims vendor not in the master vendor table

- − and more

# Orphans in Actuarial Data

- **Why is this now becoming an issue?**
  - Prevalence of Predictive Modeling and Data Mining
  - Needs to be resolved early in the data integration process

- **Data Integration for Actuarial Work**
  - Traditional Pricing and Reserving − Premium and Claims integrated at the coverage level
  - Data Mining and Predictive Modeling − Premium and Claims integrated at the policy level

# What causes orphans?

- Policy not booked, but premium and/or claims being paid.

- Endorsements

- Changing coverages

- Quotes with producers not in Agent Management System

- Manual data entry

- Differing filter definitions

# How do we test for Orphans?

- SQL − Right Outer Join (Child table on the right) with a where clause (test that record is not in Parent table)

Parent                    Child
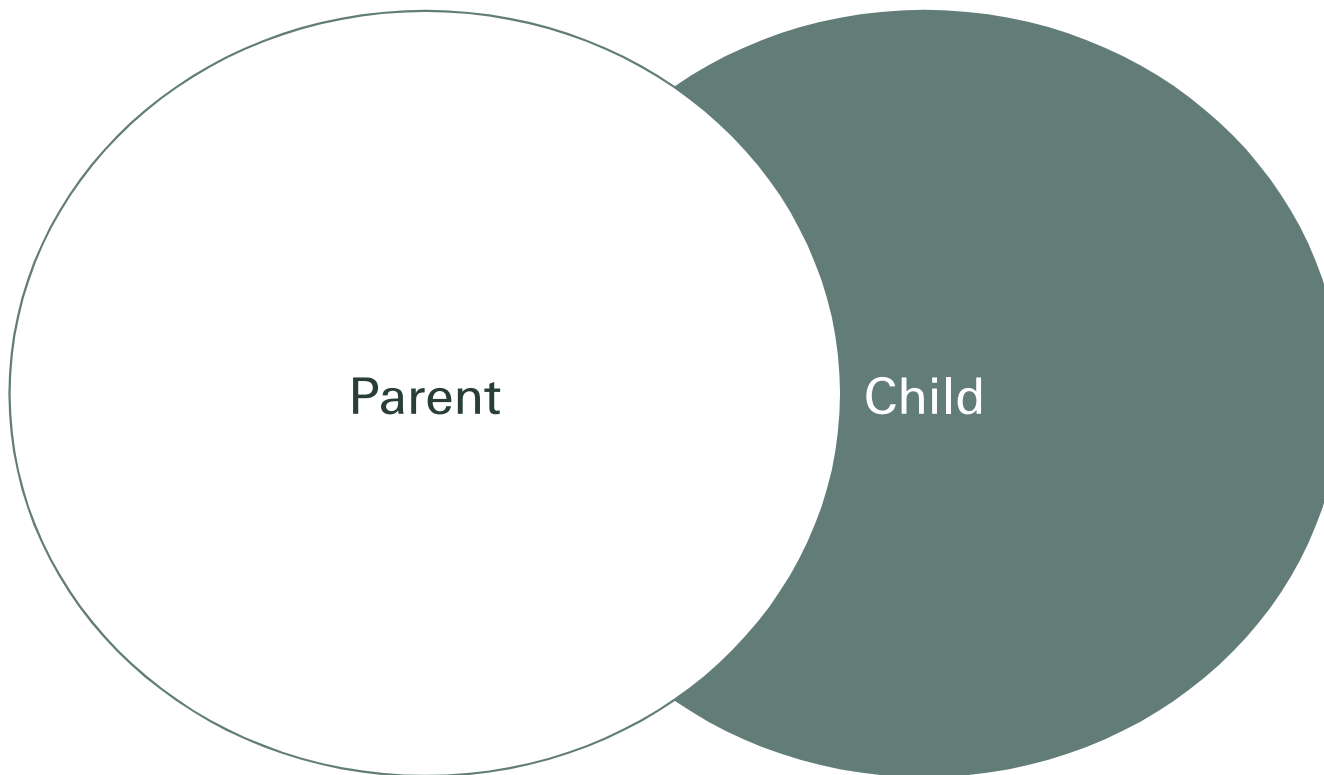
# How do we fix orphans?

- **Fix Process**
  - Systems
    - Referential Integrity
    - Underwriting, Claims and Accounting processes
  - Data Entry
    - Edits
    - Lookup Tables
    - Automate
    - Data Quality Rules

# How do we fix orphans?

- **Cleanse Data**
  - Investigate orphan records to find where they should have belonged.
    - Match policy numbers, effective dates, policy names, claimant names or any other fields that are available.
  - If you can find a match, preferred method of cleaning is to clean in source system.  Otherwise, clean within dataset.
  - If you are unable to find matches, sometimes it is best to remove from your dataset for analysis.
  - If orphan occurs because of different filtering definitions for policy vs. claims tables, then you could filter policy table first and then merge with claims table.

Questions ?

Swiss Re

# Legal notice