

Itemised Bill (IB) Table Patterns – Standard Documentation

In multi-page inpatient itemised bills, tables may appear in different structural forms. To simplify classification and automation.

Type A – Fully Structured Tables (Headers on All Pages)

Characteristics:

- Every page contains a complete and consistent header.
- Column alignment is preserved across all pages.
- Minimal OCR or formatting noise.
- Page structure is grid-like, highly regular, and predictable.

Key Notes:

- Easiest type to process automatically.
 - Common in digitally generated hospital bills.
 - high accuracy in direct extraction.
-

Type B – Continuation Tables (Headers on First Page Only)

Characteristics:

- First page shows full header.
- Subsequent pages contain rows only (no header lines).
- Sometimes marked with words like “*continued*”, “*carried forward*”, or page numbering (*Page 2 of 5*).
- Rows are still aligned in multiple columns.

Key Notes:

- Requires carrying forward headers from Page 1.

Type C – Inconsistent Headers Across Pages

Characteristics:

- Headers exist but change between pages.
- Variations may include:
 - Different wording (*Description* → *Service*).
 - Different column count (*Page 1 has 4 columns, Page 2 has 5*).
 - Different ordering of columns (*Amount before Date* on one page).
- The table remains structured, but schema is unstable.

Key Notes:

- Requires reconciliation across pages.

Type D – Noisy / Corrupted Tables

Characteristics:

- **Headers missing**, broken, or unreadable.
- OCR/scan issues introduce merged cells, uneven spacing, or fragmented text.
- Rows may not align vertically → table structure is unreliable.
- In some cases, headers may partially appear but be corrupted.

Key Notes:

- Requires additional cleaning or manual review.

Type E – Hybrid Tables (Mixed Patterns)

Characteristics:

- Document contains a **combination of multiple patterns** across different pages.
- Example:
 - Page 1 follows Type A.
 - Page 2 follows Type B.
 - Page 3 is corrupted like Type D.

Key Notes:

- Most complex case → needs page-wise classification before merging.



Rules for Each Pattern

Pattern A – Fully Structured Tables

Detection Rules:

1. Every page contains a **header row** (e.g., *Description, Qty, Rate, Amount*).
2. Headers are **consistent** (same names, same order) across all pages.
3. Each row is **aligned with its respective column**.
4. Columns have **clear boundaries** (no merged/missing columns).
5. Totals are usually given at the end of the table.

Description Rules:

- Cleanest, most structured pattern.
- Usually digital, not scanned.
- Each table behaves like an Excel sheet.

Edge Cases:

- Sometimes hospitals add **extra notes** at the bottom (e.g., “*All charges inclusive of tax*”). These should not be confused with table rows.

Pattern B – Continuation Tables

Detection Rules:

1. First page has a **full header row**.
2. Subsequent pages contain **only rows, no headers**.
3. Page markers like “*Continued...*” or page numbers may appear.
4. Columns remain **consistent** across pages (same number/order as Page 1).
5. Total usually appears at the very end, not on every page.

Description Rules:

- Long bills are split across multiple pages.
- Only the **first page sets the schema**, and later pages follow it.

Edge Cases:

- Sometimes continuation pages have **partial headers** (e.g., only “Description | Amount”).
 - Need to be careful that page numbers or footnotes are not captured as table rows.
-

Pattern C – Inconsistent Headers

Detection Rules:

1. Each page has a header, **but header names/order vary**.
 - Example: Page 1 → “Description, Qty, Amount”; Page 2 → “Service, Amount, Date”.
2. Columns may be **added/removed** on different pages.
3. Column naming may be **different words but same meaning** (e.g., “Service” vs. “Procedure”).
4. Data rows are aligned, but the schema keeps shifting.

Description Rules:

- Still structured as a table, but **unstable schema**.
- Happens when bills are generated from **different hospital departments**.

Edge Cases:

- Sometimes inconsistent headers come only from a **minor variation in spacing** (e.g., “Qty” vs. “Quantity”).
 - Need to distinguish between **true inconsistency** and **spelling differences**.
-

Pattern D – Noisy / Corrupted Tables

Detection Rules:

1. Headers may be **missing, merged, or broken**.
 - Example: “Rev Code Date” instead of two separate columns.
2. Rows may be **misaligned** (data spills across wrong columns).
3. OCR/scan errors may merge text or split words incorrectly.
4. Column count may vary between rows.
5. Some data may appear as **floating text** instead of inside the table.

Description Rules:

- The most difficult pattern.
- Usually caused by **scanning + OCR issues**.
- Not fully structured — needs cleaning and reconstruction.

Edge Cases:

- Sometimes only **part of the bill** is corrupted (rest may be Pattern A or B).

- Must confirm whether corruption is **global (all pages)** or **local (few pages only)**.
-

Pattern E – Hybrid Tables

Detection Rules:

1. Different pages of the same document follow **different patterns** (A–D).
 - Example: Page 1 = Fully Structured (A), Page 2 = Continuation (B), Page 3 = Noisy (D).
2. Some pages may have **multiple mini-tables** with different schemas.
3. Same document may include **both digital + scanned pages**.

Description Rules:

- A **mixture** of two or more patterns.
- Common in **insurance/compiled bills** (where hospital, lab, pharmacy records are merged).

Edge Cases:

- Sometimes 90% of the document is one pattern (e.g., A), but 1–2 pages are inconsistent.
 - Should be labeled as **Hybrid** if more than 1 pattern exists in the same document.
-

Summary of Rules

Pattern

How to Detect

Key Challenge

| | | |
|---------------------------------|---|----------------------------------|
| A – Fully Structured | Headers on every page, consistent schema | Very clean, easy |
| B – Continuation | Headers only on first page, rows continue | Must carry schema forward |
| C – Inconsistent Headers | Headers change (names/order/columns) | Schema unstable |
| D – Noisy/Corrupted | Headers missing/broken, OCR issues | Cleaning & reconstruction needed |
| E – Hybrid | Mix of A–D in one file | Page-by-page handling |