



# Comparison of PyMuPDF + Unstructured Vs Adobe

Analysis & Examples

Feb 13<sup>th</sup>, 2024

# Agenda Slide

## Background and Objective

- Comparison Adobe vs Combination of PyMUPDF and Unstructured
- Application output Comparison
  - Response Comparison
  - Accurate Table Parsing
- Parser output Comparison
  - Paragraph Order and Segmentation Comparison
  - Parser Output Comparison – Tables

# Background and Objective

## Background

- Application and Use of PDF parser
  - PDF parsers are employed to extract data from PDF files, transforming it into a Text format that can be ingested into a vector database. This data can then be utilized to retrieve pertinent information in response to user queries.
- Challenges
  - The extraction of data from PDFs can pose significant challenges due to the unstructured nature and diverse styles employed in different files.
  - Accurate classification of images, tables, and paragraph continuity, as well as extracting text from scanned PDFs (OCR), can be particularly difficult.

## Objective

- The solution lies in utilizing PDF parsers that are trained using machine learning. These parsers are capable of understanding the relative positioning and reading order across columns and page breaks, thereby enhancing the accuracy and efficiency of data extraction.
- We conducted a comparison analysis across multiple PDF parsers and tested them on selected PDFs using various parameters. The current analysis specifically contrasts the performance of Adobe PDF parser and combination of PyMUPDF and Unstructured . The objective is to determine the most suitable PDF parsing approach in alignment with data security policies.

# Comparison Adobe vs Combination of PyMuPDF and Unstructured

Parameters	Adobe PDF Parser	Combination of two packages		
		PyMUPDF + Unstructured Hi Res	PyMuPDF	Unstructured – Hi Res (Detectron)
Scanned document	Yes	Yes <i>Unstructured – Hi Res</i>	No	Yes
paragraph order	Most Precise – (Numbered, Order Maintained)	Above Average <i>PyMUPDF</i>	Above Average (Not Numbered, Header, footnotes order not maintained, Impacted by Tables, Other Blocks)	Average (Not Numbered, Header, footnotes order not maintained, Impacted by Tables, Images & Other Blocks)
Columns	Able to read complex docs (3 columns)	Able to read but may loose order & content in some places. <i>PyMUPDF</i>	Able to read complex docs (3 columns) but needs cleaning (example extra \n)	Able to read but may loose order & content in some places.
Tagging	Most Precise (Images, Tables, P, H1, H2, Footnotes....etc)	Above Average Tagging (Deployed User defined tagging model for tagging and cleaning leveraging font information) <i>PyMUPDF</i>	No Explicit Tags (text info eg, font size & name which can utilized for user defined tagging model) - Creates obj for tables	Below Average Tagging (No Image tags, wrong title, header tags)
Tables	Most Precise	Good (Missing content, HTML format) <i>Unstructured – Hi Res</i>	Poor (Missing content, Reading normal text as table, loosing uniformity)	Good (Missing content, HTML format)
Images/ Graphs	Yes	Need to explore <i>PyMUPDF</i>	Yes (only RGB & Grayscale images)	Need to explore
Time Complexity	3-6 sec per page (depends on complexity)	Relatively higher than Adobe and individual packages. (Need to explore & utilize parallel processing)	1 sec for 1-2 pages (depends on complexity)	Relatively higher (Need to explore & utilize parallel processing)

\* Average refer to metrics in terms of Quality

# Application output Comparison

Comparing the Application query response for Adobe and Combination of PyMUPDF and Unstructured

Question : What are the therapies for moderate active diseases

Original PDF table

TABLE 1. Treatments for thyroid eye disease			
Therapy	Mode of Action	Pros and Cons	Common Doses
Mild active disease			
Topical solutions			
Artificial tears	Maintain tear film	Rapid action and minimal side effects	
Glucocorticoids	Reduce inflammation	Rapid action and minimal side effects	
Avoidance of wind, light, dust, smoke	Reduces ocular surface irritation		
Elevation of head during sleep			
Avoidance of eye cosmetics		Benefits not yet confirmed	
Selenium	Uncertain	Benefits not yet confirmed	
Moderate or severe active disease			
Systemic glucocorticoids			
Oral	Reduce inflammation and orbital congestion	Hyperglycemia, hypertension, and osteoporosis	Up to 100 mg of oral prednisone daily, followed by tapering of the dose
Intravenous	Reduce inflammation and orbital congestion	Rapid onset of anti-inflammatory effect, fewer side effects than oral delivery, and liver damage on rare occasions	Methylprednisolone, 500 mg/week for 6 weeks, followed by 250 mg/week for 6 weeks
Orbital irradiation	Reduces inflammation	Can induce retinopathy	2 Gy daily for 2 weeks (20 Gy total)
B-cell depletion*	Reduces autoreactive B-cells	Very expensive; risks of infection, cancer, and allergic reaction	Two 1,000 mg doses of intravenous rituximab 2 weeks apart
Emergency orbital decompression†	Reduces orbital volume	Surgical procedure with inherent risks, such as postoperative diplopia	
Stable disease (inactive)			
Orbital decompression (fat removal)	Reduces orbital volume	Postoperative diplopia and pain	
Bony decompression of the lateral and medial walls	Reduces proptosis by enlarging orbital space	Postoperative diplopia, pain, sinus bleeding, and cerebrospinal fluid leak	
Strabismus repair	Improves eye alignment and reduces diplopia	Surgical procedure with inherent risks	
Eyelid repair	Improves appearance, reduces lagophthalmos, and improves function	Surgical procedure with inherent risks, including orbital hemorrhage	

Correct response

TED moderate active diseases

TED stable diseases

Unstructured + PyMuPDF Parser

What are the therapies for Moderate active disease

Therapies for Moderate active disease

1

• Systemic glucocorticoids (Oral): Reduce inflammation and orbital congestion. Common doses include up to 100 mg of oral prednisone daily, followed by tapering of the dose.

2

• Intravenous Orbital irradiation: Reduces inflammation and orbital congestion. Common dose is Methylprednisolone, 500 mg/week for 6 weeks, followed by 250 mg/week for 6 weeks.

3

• B-cell depletion\*: Reduces autoreactive B-cells. Common dose is two 1,000 mg doses of intravenous rituximab 2 weeks apart.

4

• Emergency orbital decompression†: Reduces orbital volume. It is a surgical procedure with inherent risks.

5

• Orbital decompression (fat removal): Reduces orbital volume. Common side effects include postoperative diplopia and pain.

6

• Bony decompression of the lateral and medial walls: Reduces proptosis by enlarging orbital space. Common side effects include postoperative diplopia, pain, sinus bleeding, and cerebrospinal fluid leak.

7

• Strabismus repair: Improves eye alignment and reduces diplopia. It is a surgical procedure with inherent risks.

8

• Eyelid repair: Improves appearance, reduces lagophthalmos, and improves function. It is a surgical procedure with inherent risks, including orbital hemorrhage.

Insights

- **Wrong Information** in response & **missed therapy details.** (highlighted in red circles)
- For Moderate active disease, only first 5 points (marked in blue) should come while last 4 therapies are also included in response.
- It shows that parser missed some text & parsed with wrong order for some rows.

Adobe Parser



What are therapies for Moderate active diseases?

For moderate active disease, the following therapies can be considered:

1

- Systemic glucocorticoids: These can help reduce inflammation and orbital congestion. Common doses include up to 100 mg of oral prednisone daily, followed by tapering of the dose.

2

- Intravenous glucocorticoids: These can also reduce inflammation and orbital congestion. They have a rapid onset of anti-inflammatory effect and fewer side effects than oral delivery. A common dose is methylprednisolone, 500 mg/week for 6 weeks, followed by 250 mg/week for 6 weeks.

3

- Orbital irradiation: This therapy can reduce inflammation but can induce retinopathy. The recommended dose is 2 Gy daily for 2 weeks (20 Gy total).

4

- B-cell depletion: This therapy reduces autoreactive B-cells but is very expensive and carries risks of infection, cancer, and allergic reaction. The recommended regimen is two 1,000 mg doses of intravenous rituximab 2 weeks apart.

5

- Emergency orbital decompression: This surgical procedure can reduce orbital volume but has inherent risks, such as postoperative diplopia.

Please note that these are general recommendations and the specific treatment plan should be determined by a healthcare professional based on individual patient factors and preferences.

Sources

Insights

- **Correct Response**

# Accurate Table Parsing

Original PDF table

Case	Gender	Age	Smoker?	Duration of TED (months)	CAS
Non-Inflammatory TED 1	M	43	Yes (20 years)	9	1
Non-Inflammatory TED 2	M	70	No	18	1
Non-Inflammatory TED 3	F	47	No	12	1
Non-Inflammatory TED 4	F	33	No	9	1
Non-Inflammatory TED 5	F	57	No	108	1
Inflammatory TED 1	M	29	Yes (8 years)	6	5
Inflammatory TED 2	M	35	Yes (10 years)	14	4
Inflammatory TED 3	F	50	No	88	3
Inflammatory TED 4	F	49	No	15	3
Inflammatory TED 5	M	53	Yes (30 years)	12	3
Control 1	M	48	No	—	—
Control 2	F	63	No	—	—
Control 3	M	47	No	—	—
Control 4	F	42	No	—	—
Control 5	M	53	No	—	—



Unstructured + PyMuPDF Parser



What is Duration of TED for Non-Inflammatory TED 1 case with age 43?



The duration of TED for the Non-Inflammatory TED 1 case with age 43 is 9 months.



Adobe Parser

What is Duration of TED for Non-Inflammatory TED 1 case with age 43?

NA



The duration of TED for the Non-Inflammatory TED 1 case with age 43 is 9 months.

Sources



Copy

Regenerate

Like

Leave a comment

# Parser output Comparison

Comparing the parser generated output of Adobe and Combination of PyMUPDF and Unstructured



# Paragraph Order and Segmentation Comparison

Unstructured + PyMuPDF Parser

Vs

Adobe Parser



Key points

Proptosis and Diplopia Response in Moderate to Severe Thyroid Eye Disease

Original Investigation Research

**T**h thyroid eye disease (TED), or Graves ophthalmopathy, is an autoimmune disorder characterized by progressive inflammation and damage to orbital and ocular tissues.<sup>1,2</sup> Age-adjusted prevalence in the US is estimated at 0.25%.<sup>3</sup> Thyroid eye disease causes expansion of retro-orbital fat and extraocular muscle, thought to be mediated primarily by the upregulation of the insulin like growth factor 1 receptor on orbital fibroblasts.<sup>3</sup> Patients may develop considerable disfiguring facial changes owing to proptosis, disabling diplopia, and in severe cases, vision loss.<sup>1</sup>

Currently there are limited noninvasive treatment options that improve proptosis and diplopia. The European Group on Graves' Orbitopathy (EUGO) lines recommend a cumulative dosage of 4.5 to 10 mg of intravenous methylprednisolone (IVMP) over 12 weeks for most patients with moderate to severe active TED.<sup>4</sup> Although data demonstrate that IVMP is associated with reduced inflammation, the dose, timing of administration, and duration of therapy vary in the literature, making it challenging to compare the clinical results, particularly on the progressive outcomes of proptosis and diplopia. A 2-mm reduction in proptosis and a 1-grade improvement in diplopia have been considered clinically meaningful in prior TED clinical trials.

On January 21, 2020, teprotumumab became the first US Food and Drug Administration-approved treatment for TED.<sup>5,6</sup> Teprotumumab, a fully human, monoclonal antibody, inhibits insulin like growth factor 1 receptor activity and reduces downstream pathogenic signaling in TED. A total of 2 placebo-controlled, double-masked, randomized clinical trials (RCTs) of patients with moderate to severe TED demonstrated that teprotumumab was associated with clinically significant reductions in inflammation, proptosis, and diplopia over 24 weeks.<sup>7,8</sup>

To our knowledge, there are currently no studies directly comparing the efficacy of the most recommended dose of IVMP with teprotumumab or placebo; as such, matching-adjusted indirect comparisons (MAICs) simulating direct comparisons between treatments can be used to estimate comparative treatment effects. The objectives of this study are to (1) to evaluate improvements in proptosis and diplopia with the most recommended treatment regimen of IVMP as reported in the literature and (2) to compare these results with teprotumumab and placebo in patients with moderate to severe active TED using MAICs.

Methods

Patients Receiving Teprotumumab and Placebo

Data sources included deidentified patient-level data for teprotumumab or placebo from the phase 2 (NCT01868997) and 3 (NCT03298867) trials and published aggregate-level data for IVMP (4.5-5 g over 12 weeks). Data for patients receiving teprotumumab or placebo were obtained from 2 published trials: a phase 2 trial that included 43 patients and 45 patients in the teprotumumab and placebo groups, respectively, and a phase 3 trial that included 41 patients and 42 patients in the teprotumumab and placebo groups.<sup>7,8</sup> Given the similar

Key Points

**Question** Is teprotumumab more efficacious than intravenous methylprednisolone (IVMP) for proptosis and diplopia?

**Findings** This meta-analysis and matching-adjusted indirect comparison showed an association with small improvements in proptosis from baseline for IVMP vs placebo (-0.16 mm); associated proptosis improvements were statistically significantly greater with teprotumumab vs IVMP treatment difference, -2.31 mm). For diplopia response, IVMP was not favored over placebo while teprotumumab was favored over IVMP.

**Meaning** Improvements in proptosis and diplopia with IVMP vs placebo may be small/not clinically relevant; in this meta-analysis, teprotumumab was associated with greater improvements in proptosis and diplopia vs IVMP, but clinical trials are needed to confirm the clinical relevance of this finding.

Para 11

Inclusion and exclusion criteria, data were pooled to obtain treatment arms with 84 randomized patients and 87 randomized patients for teprotumumab and placebo, respectively.

Literature Review for IVMP

A literature review was conducted to identify existing published literature assessing the most commonly recommended dose of IVMP among patients with moderate to severe active TED.<sup>9</sup> PubMed and Embase were searched for relevant RCTs and observational studies from database inception to date of search (October 5, 2020) using a search strategy that included key terms and controlled vocabulary (eg, "intravenous steroid," "Graves' orbitopathy," "thyroid eye disease," "Graves' ophthalmopathy") (search strategy presented in eAppendix 1 in the Supplement). Results were filtered to include only studies conducted in humans. Regular alerts were established to capture any recent studies until April 1, 2021.

Screening and Selection Criteria

Study inclusion was based on PICOS (population, intervention, comparator, outcomes, and study design) criteria established a priori. Briefly, only studies including patients with moderate to severe active TED receiving treatment with IVMP at a dosage of 4.5 g to 5 g over 12 weeks and reporting at least 1 of the 2 outcomes of interest (ie, change from baseline in proptosis in millimeters and/or Bahn-Gorman diplopia score) were included.<sup>10</sup> Two reviewers (R.A.Q. and R.B.) independently reviewed each title and abstract to identify eligible studies. Full texts of eligible studies were also examined for inclusion criteria and then reviewed to catalog the results.

Data Extraction

Data were extracted by a single reviewer (R.A.Q.) and verified for accuracy by a second reviewer (R.B.). Data extraction was completed using a standardized form and included study characteristics (eg, authors, study design), eligibility criteria (ie, inclusion and exclusion criteria), patient baseline characteristics (eg, sample sizes, sex, age, smoking status), and trial outcomes (eg, change from baseline in proptosis).

RCTs and observational studies from database inception to date of search (October 5, 2020) using a search strategy that included key terms and controlled vocabulary (eg, "intravenous steroid," "Graves' orbitopathy," "thyroid eye disease," "Graves' ophthalmopathy") (search strategy presented in eAppendix 1 in the Supplement). Results were filtered to include only studies conducted in humans. Regular alerts were established to capture any recent studies until April 1, 2021.

8, 0  
311.78192138671875,  
488.15826416015625,  
539.4005126953125,  
622.5857543945312,  
"Screening and Selection Criteria"

Study inclusion was based on PICOS (population, intervention, comparator, outcomes, and study design) criteria established a priori. Briefly, only studies including patients with moderate to severe active TED receiving treatment with IVMP at a dosage of 4.5 g to 5 g over 12 weeks and reporting at least 1 of the 2 outcomes of interest (ie, change from baseline in proptosis in millimeters and/or Bahn-Gorman diplopia score) were included. Two reviewers (R.A.Q. and R.B.) independently reviewed each title and abstract to identify eligible studies. Full texts of eligible studies were also examined for inclusion criteria and then reviewed to catalog the results.

Last para

311.7825927734375,  
637.3286743164062,  
539.3622436523438,  
725.8577278507812,  
"Data Extraction"

Data were extracted by a single reviewer (R.A.Q.) and verified for accuracy by a second reviewer (R.B.). Data extraction was completed using a standardized form and included study characteristics (eg, authors, study design), eligibility criteria (ie, inclusion and exclusion criteria), patient baseline characteristics (eg, sample sizes, sex, age, smoking status), and trial outcomes (eg, change from baseline in proptosis).

Key points

323.7561950683594,  
91.281982421875,  
515.3381958007812,  
108.74320983886719,  
"Question Is teprotumumab more efficacious than intravenous methylprednisolone (IVMP) for proptosis and diplopia?"

12,  
323.7561950683594,  
91.281982421875,  
515.3381958007812,  
108.74320983886719,  
"Question Is teprotumumab more efficacious than intravenous methylprednisolone (IVMP) for proptosis and diplopia?"

## Insights

- Tables and images, similar to key points, are typically placed at the end of the page, potentially disrupting the following sequence.

## JSON OUTPUT

```
> P[7]: Currently there are limited noninvasive treatment opti...
  Path:  "//Document/P[7]"

> P[8]: On January 21, 2020, teprotumumab became the first US ...
  Path:  "//Document/P[8]"

> P[9]: To our knowledge, there are currently no studies direc...
  Path:  "//Document/P[9]"

> H1: Methods
  Path:  "//Document/H1"

2: Patients Receiving Teprotumumab and Placebo
  Path:  "//Document/H2"

> P[10]: Data sources included deidentified patient-level data...
  Path:  "//Document/P[10]"

> P:
  Bounds: [323.76, 706.48, 367.20, 718.19]
  > Font:
  HasClip: false
  Lang: "en"
  Page: 1
  Path:  "//Document/Aside[2]/P"
  Text: "Key Points"
  TextSize: 8.98
  > attributes:

> P[2]: Question Is teprotumumab more efficacious than intrave...
  Path:  "//Document/Aside[2]/P[2]"

> P[3]: Findings This meta-analysis and matching-adjusted indi...
  Path:  "//Document/Aside[2]/P[3]"

> P[4]: Meaning Improvements in proptosis and diplopia with IV...
  Path:  "//Document/Aside[2]/P[4]"

> P[11]: inclusion and exclusion criteria, data were pooled to...
  Path:  "//Document/P[11]"

> H2[2]: Literature Review for IVMP
  Path:  "//Document/H2[2]"

> P[12]: A literature review was conducted to identify existin...
  Path:  "//Document/P[12]"

> H2[3]: Screening and Selection Criteria
  Path:  "//Document/H2[3]"

> P[13]: Study inclusion was based on PICOS (population, inter...
  Path:  "//Document/P[13]"
```

Para 10

Key points

Para 11

## Insights

- Ensure that the sequence of para, such as para 10 followed by para 11, is maintained without any disruption.
- Key points is tagged differently hence help to identify right sequence.

Last para



# Parser Output Comparison – Tables

## Original PDF table

TABLE 4. Summary of recommendations on usage of teprotumumab for thyroid eye disease

	Recommendations
Treatment population	Age: adults; use with in caution in postpubertal adolescents Thyroid status: any, can be started concomitantly with attempts to achieve euthyroidism TED status: progressive disease
Contraindications	TED severity: clinically significant Previous treatment of thyroid/TED: any Poorly controlled diabetics, pregnant or planning to become pregnant, nursing mothers, and prepubertal children
	Those on concomitant biologics, or those who received rituximab within 6 months Patients with inflammatory bowel disease should be treated with caution and comanaged with gastroenterologists
Dose and duration	1st infusion: 10 mg/kg (over 90 minutes) 2nd infusion: 20 mg/kg (over 90 minutes) Subsequent infusions: 20 mg/kg (over 60 minutes) Total of 8 infusions (every 3 weeks) Early discontinuation if no improvement at 4–6th infusion Additional doses if severe disease with continued improvement
Pre-infusion screen	Complete medical (including weight and BP measurement) and ophthalmic examination Baseline laboratory results: fasting blood glucose, HgA1c, LFTs, and CBC Baseline EKG
Drug monitoring	In patients with diabetes: fasting blood glucose after each of the 1st 2 infusions. Self-monitoring at least twice a day. Work in conjunction with endocrinologist.

BP, blood pressure; CBC, complete blood count; EKG, electrocardiography; HgA1c, hemoglobin A1c; LFTs, liver function tests; TED, thyroid eye disease.

## Unstructured + PyMuPDF Parser

	Recommendations
Treatment population	Age: adults; use with in caution in postpubertal adolescents thyroid status: any, can be started concomitantly with attempts to achieve euthyroidism ED status: progressive disease ED severity: clinically significant those on concomitant biologics, or those who received rituximab within 6 months Patients with inflammatory bowel disease should be treated with caution and comanaged with gastroenterologists
Dose and duration	1st infusion: 10 mg/kg (over 90 minutes) 2nd infusion: 20 mg/kg (over 90 minutes) Subsequent infusions: 20 mg/kg (over 60 minutes) otal of 8 infusions (every 3 weeks) Early discontinuation if no improvement at 4–6th infusion Additional doses if severe disease with continued improvement
Pre-infusion screen	Complete medical (including weight and BP measurement) and ophthalmic examination Baseline laboratory results: fasting blood glucose, HgA1c, LFTs, and CBC Baseline EKG
Drug monitoring	In patients with diabetes: fasting blood glucose after each of the 1st 2 infusions. Self-monitoring at least twice a day. Work in conjunction with endocrinologist.

### Insights

#### Inaccurate table data Reading

- Missed few rows (marked in red box in Original PDF Table)
- Some words are incomplete (highlighted in yellow in the above table).

Vs

## Adobe Parser



	Recommendations
Treatment population	Age: adults; use with in caution in postpubertal adolescents Thyroid status: any, can be started concomitantly with attempts to achieve euthyroidism TED status: progressive disease TED severity: clinically significant Previous treatment of thyroid/TED: any
Contraindications	Poorly controlled diabetics, pregnant or planning to become pregnant, nursing mothers, and prepubertal children Those on concomitant biologics, or those who received rituximab within 6 months Patients with inflammatory bowel disease should be treated with caution and comanaged with gastroenterologists
Dose and duration	1st infusion: 10 mg/kg (over 90 minutes) 2nd infusion: 20 mg/kg (over 90 minutes) Subsequent infusions: 20 mg/kg (over 60 minutes) Total of 8 infusions (every 3 weeks) Early discontinuation if no improvement at 4–6th infusion Additional doses if severe disease with continued improvement
Pre-infusion screen	Complete medical (including weight and BP measurement) and ophthalmic examination Baseline laboratory results: fasting blood glucose, HgA1c, LFTs, and CBC Baseline EKG
Drug monitoring	In patients with diabetes: fasting blood glucose after each of the 1st 2 infusions. Self-monitoring at least twice a day. Work in conjunction with endocrinologist.

### Insights

- Accurate table structure identification.
- Returns parsed output in csv format.

# Parser Output Comparison – Tables

## Original PDF table

**TABLE 3.** Adverse events from combined Phase 2 and Phase 3 trials

Adverse Event, N (%)	Teprotumumab (N = 85)	Placebo (N = 86)
Muscle spasm	21 (25%)	6 (7%)
Nausea	14 (17%)	8 (9%)
Alopecia	11 (13%)	7 (8%)
Diarrhea	10 (12%)	7 (8%)
Fatigue	10 (12%)	6 (7%)
Hyperglycemia	8 (10%)	1 (1%)
Hearing impairment*	8 (10%)	0
Dysgeusia	7 (8%)	0
Headache	7 (8%)	6 (7%)
Dry skin	7 (8%)	0

\*Hearing impairment includes deafness, Eustachian tube dysfunction, hyperacusis, hypoacusis, and autophony.

## Unstructured + PyMuPDF Parser

**Adverse Event, N (%) Teprotumumab (N = 85) Placebo (N = 86)**

Muscle spasm	21 (25%)	
Nausea	14 (17%)	
Alopecia	11 (13%)	
Diarrhea	10 (12%)	
Fatigue	10 (12%)	
Hyperglycemia	8 (10%)	
Hearing impairment*	8 (10%)	
Dysgeusia	7 (8%)	
Headache	7 (8%)	
Dry skin	7 (8%)	

### Insights

1. Inaccurate table data identification.
2. Missed column values (marked in red box)

Vs

## Adobe Parser



Adverse Event, N (%)	Teprotumumab (N = 85)	Placebo (N = 86)
Muscle spasm	21 (25%)	6 (7%)
Nausea	14 (17%)	8 (9%)
Alopecia	11 (13%)	7 (8%)
Diarrhea	10 (12%)	7 (8%)
Fatigue	10 (12%)	6 (7%)
Hyperglycemia	8 (10%)	1 (1%)
Hearing impairment*	8 (10%)	0
Dysgeusia	7 (8%)	0
Headache	7 (8%)	6 (7%)
Dry skin	7 (8%)	0

### Insights

1. Accurate table data identification.
2. Returns parsed output in csv format.

## Original PDF table

**TABLE 1.** Treatments for thyroid eye disease

Therapy	Mode of Action	Pros and Cons	Common Doses
<b>Mild active disease</b>			
<b>Topical solutions</b>			
Artificial tears	Maintain tear film	Rapid action and minimal side effects	
Glucocorticoids	Reduce inflammation	Rapid action and minimal side effects	
Avoidance of wind, light, dust, and smoke	Reduces ocular surface desiccation and reduces irritation		
Elevation of head during sleep	Reduces orbital congestion		
Avoidance of eye cosmetics	Reduces irritation	Benefits not yet confirmed	
Selenium	Uncertain	Benefits not yet confirmed	
<b>Moderate or severe active disease</b>			
<b>Systemic</b>			
glucocorticoids Oral	Reduce inflammation and orbital congestion	Hyperglycemia, hypertension, and osteoporosis	Up to 100 mg of oral prednisone daily, followed by tapering of the dose
Intravenous	Reduce inflammation and orbital congestion	Rapid onset of anti-inflammatory effect, fewer side effects than oral delivery, and liver damage on rare occasions	Methylprednisolone, 500 mg/week for 6 weeks, followed by 250 mg/week for 6 weeks 2 for 2 weeks
Orbital irradiation	Reduces inflammation	Can induce retinopathy	2 Gy daily for 2 weeks (20 Gy total)
B-cell depletion*	Reduces autoreactive B-cells	Very expensive; risks of infection, cancer, and allergic reaction	Two 1,000 mg doses of intravenous rituximab 2 weeks apart
Emergency orbital decompression†	Reduces orbital volume	Surgical procedure with inherent risks, such as postoperative diplopia	
<b>Stable disease (inactive)</b>			
Orbital decompression (fat removal)	Reduces orbital volume	Postoperative diplopia and pain	
Bony decompression of the lateral and medial walls	Reduces proptosis by enlarging orbital space	Postoperative diplopia, pain, sinus bleeding, and cerebrospinal fluid leak	
Strabismus repair	Improves eye alignment and reduces diplopia	Surgical procedure with inherent risks	
Eyelid repair	Improves appearance, reduces lagophthalmos, and improves function	Surgical procedure with inherent risks, including orbital hemorrhage	

Reproduced with permission from (2).

\*B-cell depletion with the use of rituximab is not approved by the Food and Drug Administration for this indication.

†Emergency orbital decompression is indicated for optic neuropathy or severe corneal exposure.

## Insights

Inaccurate table data reading & structure identification -

1. Combined 2 rows information in one. (highlighted in yellow).
2. Wrong order of text.

## Unstructured + PyMuPDF Parser

Therapy	Mode of Action	Pros and Cons	Common Doses
<b>Mild active disease</b>			
Topical solutions Artificial tears	Maintain tear film	Rapid action and minimal side effects	
Glucocorticoids	Reduce inflammation	effects Rapid action and minimal side effects	
Avoidance of wind, light, dust, and smoke	Reduces ocular surface desiccation and reduces irritation		
Elevation of head during sleep	Reduces orbital congestion		
Avoidance of eye cosmetics	Reduces irritation	Benefits not yet confirmed	
Selenium	Uncertain	Benefits not yet confirmed	
<b>Moderate or severe active confirmed disease Systemic</b>			
glucocorticoids Oral	Reduce inflammation and orbital congestion	Hyperglycemia, hypertension, and osteoporosis	Up to 100 mg of oral prednisone daily, followed by tapering of the dose
Intravenous	Reduce inflammation and orbital congestion	Rapid onset of anti-inflammatory effect, fewer side effects than oral delivery, and liver damage on rare occasions	Methylprednisolone, 500 mg/week for 6 weeks, followed by 250 mg/week for 6 weeks 2 for 2 weeks
B-cell depletion*	Reduces autoreactive B-cells	Very expensive; risks of infection, cancer, and allergic reaction	Gy daily (20 Gy total) Two 1,000 doses of mg intravenous rituximab 2 weeks apart
Emergency orbital decompression†	Reduces orbital volume	Surgical procedure with inherent risks, such as postoperative diplopia	
<b>Stable disease (inactive)</b>			
Orbital decompression (fat removal)	Reduces orbital volume	Postoperative diplopia and pain	
Bony decompression of the lateral and medial walls	Reduces proptosis by enlarging orbital space	Postoperative diplopia, pain, sinus bleeding, and cerebrospinal fluid leak	
Strabismus repair	Improves eye alignment and reduces diplopia	Surgical procedure with inherent risks	
Eyelid repair	Improves appearance, reduces lagophthalmos, and improves function	Surgical procedure with inherent risks, including orbital hemorrhage	

## Adobe Parser



Vs

Therapy	Mode of Action	Pros and Cons	Common Doses
<b>Mild active disease</b>			
Topical solutions			
Artificial tears	Maintain tear film	Rapid action and minimal side effects	
Glucocorticoids	Reduce inflammation	Rapid action and minimal side effects	
Avoidance of wind, light, dust, and smoke	Reduces ocular surface desiccation and reduces irritation		
Elevation of head during sleep	Reduces orbital congestion		
Avoidance of eye cosmetics	Reduces irritation	Benefits not yet confirmed	
Selenium	Uncertain	Benefits not yet confirmed	
<b>Moderate or severe active disease</b>			
<b>Systemic glucocorticoids</b>			
Oral	Reduce inflammation and orbital congestion	Hyperglycemia, hypertension, and osteoporosis	Up to 100 mg of oral prednisone daily, followed by tapering of the dose
Intravenous	Reduce inflammation and orbital congestion	Rapid onset of anti-inflammatory effect, fewer side effects than oral delivery, and liver damage on rare occasions	Methylprednisolone, 500 mg/week for 6 weeks, followed by 250 mg/week for 6 weeks
Orbital irradiation	Reduces inflammation	Can induce retinopathy	2 Gy daily for 2 weeks (20 Gy total)
B-cell depletion*	Reduces autoreactive B-cells	Very expensive; risks of infection, cancer, and allergic reaction	Two 1,000 mg doses of intravenous rituximab 2 weeks apart
Emergency orbital decompression†	Reduces orbital volume	Surgical procedure with inherent risks, such as postoperative diplopia	
<b>Stable disease (inactive)</b>			
Orbital decompression (fat removal)	Reduces orbital volume	Postoperative diplopia and pain	
Bony decompression of the lateral and medial walls	Reduces proptosis by enlarging orbital space	Postoperative diplopia, pain, sinus bleeding, and cerebrospinal fluid leak	
Strabismus repair	Improves eye alignment and reduces diplopia	Surgical procedure with inherent risks	
Eyelid repair	Improves appearance, reduces lagophthalmos, and improves function	Surgical procedure with inherent risks, including orbital hemorrhage	





BOSTON NEW YORK PRINCETON PHILADELPHIA SAN FRANCISCO TORONTO GURGAON MUNICH LONDON