

Create a new extraction job

Use this for your own PDFs. Define columns, upload files, then review the extracted table.

1 Data type

2 Columns

3 Settings

4 Upload

Data type * ?

Invoice

Scientific paper

Lease agreement

Next: Define columns

Data type

Describe the type of data you want to extract.

Select a preset such as Invoice, Scientific Paper or Lease Agreement or enter your own data type to match your documents such as Earnings Report or Contract

Name each column and explain what it means

The screenshot shows a web interface for configuring columns. It features a list of column entries, each with a name field, a data type dropdown, an interpretation field, and a 'Remove' button. The entries are:

- Study design (text) | How to interpret (optional)
- name (balloon) (text) | name and company of drug eluting balloon used in th
- name (stent) (text) | name and company of drug eluting stent used in the
- age (balloon) (number) | average age of patients treated with drug eluting ball
- age (stent) (number) | average age of patients treated with drug eluting ster
- Total patients (number) | total number of enrolled patients in this study
- patients (balloon) (text) | number of patients treated with drug eluting balloon
- patients (stent) (text) | How to interpret (optional)

At the bottom, there are buttons for 'Add column', 'Save as template', 'Manage templates', 'Back', and 'Next: Settings'. Two blue arrows point from the explanatory text below to the 'patients (balloon)' and 'patients (stent)' entries.

Column Name	Data Type	How to Interpret (optional)
Study design	text	How to interpret (optional)
name (balloon)	text	name and company of drug eluting balloon used in th
name (stent)	text	name and company of drug eluting stent used in the
age (balloon)	number	average age of patients treated with drug eluting ball
age (stent)	number	average age of patients treated with drug eluting ster
Total patients	number	total number of enrolled patients in this study
patients (balloon)	text	number of patients treated with drug eluting balloon
patients (stent)	text	How to interpret (optional)

Column name

This becomes the heading in the final results table and CSV. Keep it short and reusable across files.

How to interpret (optional)

Use this when the label alone could be ambiguous. Add the exact meaning, group, timeframe, unit, or missing-value rule.

3 Select the extraction structure and document settings

Flat or Relational

Flat is recommended for most of the extractions. Flat gives one row per PDF. Relational creates multiple rows when a document contains repeated items.

PDF handling and profile

Direct PDF is recommended. Start with Auto; use Research for papers or Structured Numeric for dense tables.

Flat ?
One row per PDF. Example: an invoice PDF → one row with invoice number, total, due date.

Relational Advanced - Beta ?
Multiple rows per PDF. Example: a clinical trial PDF → one row per medical device (device name, manufacturer, risk class).

PDF handling ?
Direct PDF Recommended
Works with both normal text PDFs and scanned documents. The complete original PDF is uploaded to the selected AI provider.

► Advanced PDF options
Direct PDF keeps page-level evidence, but highlight boxes can be sparse when the PDF has little or no embedded text.

Document profile ?
Auto
Let the system choose the best evidence and highlight strategy for this job.

Structured Numeric
Best for contracts, invoices, payment estimates, and row or amount heavy documents.

Research
Best for scientific papers, journals, and narrative evidence extraction.

Leave Auto on for the first run. Override only when you want to force a numeric or research-style evidence strategy.

Rules (optional)
Examples: normalize percentages, use study endpoint table first, return Not found when unsure.

Back

Next: Upload files

Upload the PDFs and create the job

New extraction

Define fields, upload PDFs, review instructions, then run extraction.

✓ Data type

✓ Columns

✓ Settings

4 Upload

PDF files * ?


Drag & drop PDFs here, or [browse](#)
Up to 10 PDFs per job, 15 MB each

s10557-021-07172-4.pdf763 KB [Remove](#)

Avoid confidential or sensitive documents. Not HIPAA compliant. Not a GDPR compliance tool.

Back

Create extraction job

Check the upload list

Add the PDFs, then confirm the correct files appear before continuing. A small representative first run is easier to verify.

Create extraction job

Continue only after reviewing the privacy notice and confirming the files are appropriate for the service.

Review the AI generated instructions before extraction

Objective

Confirm the overall objective matches the document set and the information you intend to extract.

Field plans

For each field, check the meaning, output format, and what should happen when the value is missing.

Policy page text is not injected automatically. Only job fields, data type, rules, and setup context are used.

OBJECTIVE
Extract study design, device identities, cohort demographics, enrollment, follow-up, and mortality outcomes for drug-eluting balloon and drug-eluting stent groups from the full scientific paper.

FIELD PLANS

Study design (study_design)
Identify the formal design of the study, such as randomized controlled trial, prospective cohort, retrospective observational study, registry, or meta-analysis.
Format: Short text describing the reported study design.
Missing policy: If no explicit design is reported after scanning the full document, return null.

name (balloon) (name_balloon)
Identify the drug-eluting balloon used in the study, including manufacturer/company.
Format: Device brand or model name followed by company/manufacturer in parentheses when reported.
Missing policy: If the balloon name or company is not reported after scanning the full document, return the available component only; return null if neither is available.

name (stent) (name_stent)
Identify the drug-eluting stent used in the study, including manufacturer/company.
Format: Device brand or model name followed by company/manufacturer in parentheses when reported.
Missing policy: If the stent name or company is not reported after scanning the full document, return the available component only; return null if neither is available.

age (balloon) (age_balloon)
Extract the reported average age for patients treated with the drug-eluting balloon.
Format: Numeric mean or median age in years, preserving qualifiers such as mean or median when stated.
Missing policy: If no balloon-group average age is reported after scanning the full document, return null.

Revise, save, or approve the setup

Optional revision feedback

Describe the specific correction or additional information when the generated instructions are not quite what you want, then click “Revise instructions” and AI will revise them.

Missing policy: If the stent patient count is not reported after scanning the full document, return null.

follow up period (follow_up_period)
Extract the reported follow-up duration for outcome assessment.
Format: Numeric duration with its reported unit, such as 12 months or 1 year.
Missing policy: If no follow-up period is reported after scanning the full document, return null.

mortality (balloon) (mortality_balloon)
Extract the reported mortality outcome for the drug-eluting balloon group at the relevant follow-up.
Format: Reported numeric mortality count; retain percentage only when no count is reported.
Missing policy: If balloon-group mortality is not reported after scanning the full document, return null.

mortality (stent) (mortality_stent)
Extract the reported mortality outcome for the drug-eluting stent group at the relevant follow-up.
Format: Reported numeric mortality count; retain percentage only when no count is reported.
Missing policy: If stent-group mortality is not reported after scanning the full document, return null.

Reusable setup "Balloon vs Stent papers data extraction setup" saved.

OPTIONAL REVISION FEEDBACK

Example: Treat 'test score' as numeric value + units when available.

Save setup for reuse Revise instructions Approve and queueing...

Save for reuse then approve

- “Save setup for reuse” if you want to reuse the exact same instructions for future extraction.
- Approve only after the instructions match your extraction goal.

7 Select a result and verify its source evidence

Select a field

Choose the result you want to check from the Files & fields list to see the highlights in the pdf.

Files & fields

s10557-021-07172-4.pdf

age (balloon)
62.6±8.8 years (mean)

age (stent)
64.0±10.5 years (mean)

follow up period
12 months for clinical outcomes; 9 months for angiographic outcomes

mortality (balloon)
0 deaths at 12 months

mortality (stent)
0 deaths at 12 months

name (balloon)
Sequent® Please paclitaxel-coated balloon (B. Braun Melsungen AG, Berlin, Germany)

name (stent)
New-generation drug-eluting stents: Resolute Integrity (Medtronic), Xience Xpedition (Abbott...)

patients (balloon)
85 patients randomized/enrolled; 84 analyzed after 1 withdrawal

patients (stent)

File: s10557-021-07172-4.pdf > Field: age (balloon)

84 patients in DCB group and 79 patients in DES group were analyzed for baseline data. The demographic and baseline clinical data of the two groups were well matched. Unstable angina pectoris was the most common disease type in both effective pre-dilation was similar in both groups. The diameter of the DCB finally implanted [2.75 (2.50 to 3.00) mm] was significantly lower than that of the DES [3.00 (2.75 to 3.50) mm] ($P<0.001$), and the ratio of the

Table 1 Patient baseline characteristics

	DCB group (n=84)	DES group (n=79)	Statistical value	P value
Demographic data				
Mean age (years) ^a	62.6±8.8	64.0±10.5	1.553	0.214
Male (%)	62 (73.8)	56 (70.9)	1.175	0.676
BMI ^a	26.0±3.0	25.4±3.0	0.025	0.874
Medical history				
Smoking (%)	46 (54.8)	42 (53.2)	0.042	0.838
Diabetes (%)	16 (19.0)	23 (29.1)	2.266	0.132
Hypertension (%)	50 (59.5)	54 (68.4)	1.375	0.241
Dyslipidemia (%)	52 (61.9)	39 (49.4)	2.595	0.107
Previous CABG (%)	1 (1.2)	3 (3.8)	1.156	0.355
Previous PCI (%)	11 (13.1)	14 (17.7)	0.671	0.413
History of atrial fibrillation (%)	2 (2.4)	5 (6.3)	1.544	0.266
History of cardiac insufficiency (%)	4 (4.8)	7 (8.9)	1.087	0.297
Type of coronary heart disease			0.943	0.815
Stable angina pectoris (%)	8 (9.5)	10 (12.7)	0.407	0.523
Unstable angina pectoris (%)	58 (69.0)	56 (70.9)	0.065	0.798
NSTEMI (%)	13 (15.5)	9 (11.4)	0.582	0.446
STEMI (%)	5 (6.0)	4 (5.1)	0.062	1.000
Distribution of coronary lesions			3.456	0.178
One-vessel disease (%)	14 (16.7)	12 (15.2)	0.066	0.797
Two-vessel disease (%)	36 (42.9)	24 (30.4)	2.725	0.099
Three-vessel disease (%)	34 (40.5)	43 (54.4)	3.181	0.075

DCB, drug-coated balloon; DES, drug-eluting stents; BMI, body mass index; CABG, coronary artery bypass graft; PCI, percutaneous coronary intervention; NSTEMI, non-ST-segment elevation myocardial infarction; STEMI, ST-segment elevation myocardial infarction

^a Expressed by $\bar{x} \pm s$, while the rest of the data is expressed as number (%). The statistical result for the counted data refers to the chi-square test results.

Compare with the source

Check the extracted data against the highlighted text in the PDF.

Note: The app highlights the text or sentence that most closely matches the extracted data. In some cases, the highlight may not be exact even though the extracted data is correct.

8 Use the review actions when a result needs attention

The interface is divided into two main sections. On the left, under 'Files & fields', there is a list of fields from a PDF file 's10557-021-07172-4.pdf'. The fields include 'age (balloon)', 'age (stent)', 'follow up period', 'mortality (balloon)', 'mortality (stent)', 'name (balloon)', 'name (stent)', and 'patients (balloon)'. The 'name (balloon)' field is selected and highlighted in blue. On the right, the detailed view for the 'name (balloon)' field is shown. It includes a breadcrumb 'File: s10557-021-07172-4.pdf > Field: name (balloon)', a 'Prev field' and 'Next field' button pair, and the field value 'Sequent® Please paclitaxel-coated balloon (B. Braun Melsungen AG, Berlin, Germany)'. Below this is a blue button 'Evidence 1 - p.2'. A light blue box contains the text 'The DCBs used were paclitaxel-coated (Sequent® Please; B/Braun Melsungen AG, Berlin, Germany)' with a 'Jump to page 2' link. Below this box are three buttons: 'Flag as wrong', 'Suggest correction', and 'Add note'. At the bottom of the right section, it says 'Page 2 / 10' with 'Prev' and 'Next' buttons. A document preview is shown at the bottom, with a mouse cursor pointing to a highlighted section. The preview shows page 656 of a document titled 'Cardiovasc Drugs Ther (2022) 36:655–664'. The highlighted section is titled 'Materials and Methods' and 'Study Design and Patient Enrolment'.

Files & fields

- s10557-021-07172-4.pdf
- age (balloon)
62.6±8.8 years (mean)
- age (stent)
64.0±10.5 years (mean)
- follow up period
12 months for clinical outcomes; 9 months for angiographic outcomes
- mortality (balloon)
0 deaths at 12 months
- mortality (stent)
0 deaths at 12 months
- name (balloon)
Sequent® Please paclitaxel-coated balloon (B. Braun Melsungen AG, Berlin, Germany)
- name (stent)
New-generation drug-eluting stents: Resolute Integrity (Medtronic), Xience Xpedition (Abbott...)
- patients (balloon)
85 patients randomized/enrolled; 84 analyzed after 1 withdrawal

File: s10557-021-07172-4.pdf > Field: name (balloon)

name (balloon) Prev field Next field

Sequent® Please paclitaxel-coated balloon (B. Braun Melsungen AG, Berlin, Germany)

Evidence 1 - p.2

The DCBs used were paclitaxel-coated (Sequent® Please; B/Braun Melsungen AG, Berlin, Germany)

[Jump to page 2](#)

Flag as wrong Suggest correction Add note

Page 2 / 10 Prev Next

Highlight from layout grounding (char_interval_normalized).

656 Cardiovasc Drugs Ther (2022) 36:655–664

Materials and Methods

Study Design and Patient Enrolment

This prospective, randomized, open-label, single-center study was designed to evaluate the efficacy and safety of DCB in the treatment of coronary de novo lesions. Patients who underwent elective percutaneous coronary intervention (PCI) at Beijing Hospital were screened. The target lesions should

79; or SYNERGY; Boston Scientific Corporation, Marlborough, MA; 7/79), or rapamycin-eluting stents (Firehawk, MicroPort, Shanghai, China; 15/79). The main exclusion criteria included acute myocardial infarction (MI) within 1 week, left ventricular ejection fraction <40%, chronic total occlusion, left main disease, or multiple vessel disease with more than one lesion requiring treatment. Detailed inclusion and exclusion criteria are listed in the Supplement file. The study flowchart is shown in Fig. 1.

Evidences and Page Navigation

Use the evidence buttons when more than one supporting evidence is available. Click on “Jump to page” for faster source checking.

Flag, correct, or note

Flag a wrong value, suggest the corrected value or add a review note.

Review every column, then download the CSV

Review extraction

CSV cells that could run spreadsheet formulas are prefixed with an apostrophe. The values shown here remain unchanged.

Start by opening one value you care about. A populated result cell can show the source quote, page, and highlight.

All PDFs Search extracted values

name (balloon)	name (stent)	patients (balloon)	patients (stent)	Study design	Total patients
Sequent® Please paclitaxel-coated b...	New-generation drug-eluting stents: ...	85 patients randomized/enrolled; 84 ...	85 patients randomized/enrolled; 79 a...	Prospective, randomized, open-label,...	170 patients

Review all columns
Drag the horizontal bar to check all the extracted fields

Files & fields

mortality (balloon)
0 deaths at 12 months

mortality (stent)
0 deaths at 12 months

name (balloon)
Sequent® Please paclitaxel-coated balloon (B. Braun

File: s10557-021-07172-4.pdf > Field: Total patients

Total patients

170 patients

Evidence 1 · p.1

Evidence 2 · p.3

Evidence 3 · p.1

After proper pre-dilation, 170 patients were enrolled and randomized to the DCB and the DES groups at 1:1 ratio.

[Jump to page 1](#)

Download CSV

Download CSV

Export the data table. CSV contains identifiers and extracted values; evidence highlights remain in the app.