

Dental AI SOAP Note Benchmark

A practical comparison guide for dentists

DentScribe.AI | Version 1.0

A dental AI scribe should turn a conversation into a clinical record you can review efficiently. The test is whether the final note keeps the right findings on the right teeth, preserves your decisions, and distinguishes completed care from future plans.

DentScribe.AI Benchmark 1.0 gives you ten focused challenges, two scattered-conversation challenges, and a comparison sheet. Judge DentScribe and any other scribe by the same visible results.

What you are testing

Look for	What a good note preserves
Correct teeth and surfaces	Each condition stays attached to its tooth, surface, site, or stated region.
Complete clinical detail	Important findings, measurements, materials, and follow-up survive in the final note.
Clear SOAP structure	Patient reports, findings, assessment, and plans are easy to locate and distinguish.
Accurate treatment status	Existing, completed, recommended, considered, declined, and deferred remain distinct.
Clinical judgment	The dentist's stated diagnosis, reasoning, priorities, and uncertainty remain intact.
Precise measurements	Values keep their units, locations, and meaning; spoken corrections are respected.
Useful grouping	Shared findings can be combined without adding teeth, spreading conditions, or losing exceptions.
No invented facts	Silence does not become a normal examination, consent, diagnosis, or treatment.
Correct context	Patient statements stay attributed; another patient or a scheduling entry does not become a finding.
A usable final note	The delivered note is readable, complete, and consistent across its sections.

Read once then test

Start with Run the comparison. Cases 1–10 isolate difficult dental details. Cases F1 and F2 then test how a scribe assembles a note from remarks scattered across a real encounter. Their source transcripts were reviewed in full; selected chronological excerpts are provided for public use.

No special software or knowledge of AI is required. A dentist checks the final notes against the supplied cases. Cases 1–10 are synthetic; F1 and F2 use de-identified source excerpts. Examples test documentation, not treatment recommendations.

Run the comparison

- 1 Keep conditions equal.** Start a new encounter for each case. Use the same text and standard dental note setting. Record the product version and template. Supply no extra history or example answers.
- 2 Use one common request.** Create a structured dental SOAP note from this encounter. Preserve the supplied clinical details and their status. Do not add unstated facts. Use US Universal tooth numbering.
- 3 Keep the first result.** Save the final note before editing. Do not replace it with a better rerun. Set one time limit for every product before testing.
- 4 Mark each case.** PASS requires all expected clinical content, no invented or conflicting clinical facts, and usable organization. Otherwise mark CLINICAL ERROR, ORGANIZATION ONLY, or NO USABLE NOTE. Clinical errors take priority over organization problems.

Equivalent wording, surface-letter order, bullets, tables, and combined Assessment and Plan sections are accepted. Exact wording and proprietary markers are unnecessary. Empty headings are optional. Findings and linked plans must remain distinguishable; personal style preferences do not change the result.

Case	Challenge	DentScribe	Scribe B
1	Dense hard tissue findings		
2	Periodontal findings and linked plans		
3	Site measurements and corrections		
4	Mixed primary and permanent teeth		
5	Endodontic uncertainty and referral		
6	Allergy and completed treatment		
7	Implants and medication status		
8	Orthodontic design and actual care		
9	Soft tissue findings and conditional follow-up		
10	Conflicts and encounter boundaries		
F1	Scattered consultation details		
F2	Extractions and other-patient stories		

Product and version _____ Template _____ Test date _____
Reviewer _____ Time limit _____ Input mode _____

Report focused-case passes out of 10 and scattered-conversation passes out of 2 separately. For each set, also count clinical errors, organization-only issues, and missing outputs. Record a brief reason for every failure. Do not describe a focused-only test as the complete Benchmark 1.0.

Passing all cases demonstrates performance on this public suite, not 100% accuracy in practice. Wrong teeth, invented care, and lost allergies matter regardless of totals. Unsupported input mode: mark not tested. Compare audio-only products separately using the same case recordings.

Case 1 Dense hard tissue findings

Encounter to submit

This is an examination only. Tooth thirty-one has distal-buccal caries and a lingual fracture. A crown is recommended. A mesial-lingual-buccal restoration is being considered as an alternative if the patient cannot proceed with the crown. Monitor thirty distal, eight mesial-distal, nine mesial, thirteen distal, four distal, and three mesial. A different lower-right posterior tooth has an old leaking amalgam; its number is not identified. Another tooth has an existing crown with lingual marginal darkening; consider replacement only if decay progresses. An upper-right posterior tooth has a large amalgam and a fracture; no number is given. The missing-tooth area is from a prior orthodontic extraction. No restorative treatment was performed today.

Clinical content that must survive

- #31: DB caries; L fracture. Crown recommended; MLB restoration considered as an alternative if the patient cannot proceed with the crown.
- #30D, #8MD, #9M, #13D, #4D, #3M: Monitor.
- Different lower-right posterior tooth: Existing amalgam with leakage; tooth number unspecified.
- Tooth with existing crown: Lingual marginal darkening; replacement considered only if decay progresses.
- Upper-right posterior tooth: Large existing amalgam; fracture; tooth number unspecified.
- Missing-tooth area: Prior orthodontic extraction.
- Examination only; no restorative treatment performed today.

What this reveals

"Several teeth have decay and fractures; crowns and fillings discussed" loses the surface map, the monitoring list, and the conditions attached to treatment.

The grouped monitoring line is concise because every tooth retains its own surfaces. The #31 line keeps caries on DB and the fracture on L. Writing "#31BDL: caries and fracture" would obscure that distinction in this particular encounter.

Fail examples. Assigning a number to an unidentified tooth; changing the watch list into a continuous tooth range; recording the crown as completed; dropping the alternative restoration or its condition.

Pass check. Every listed finding and decision is present and correctly linked. Clearly separated tooth lines are also acceptable.

Case 2 Periodontal findings and linked plans

Encounter to submit

There is at least four millimeters of alveolar bone loss throughout the dentition and generalized gingival recession with exposed root surfaces. Areas have four-to-five-millimeter pockets, with limited bleeding and inflammation. Scaling and root planing is planned. The clinician identifies inadequate interdental cleaning and forceful electric-toothbrush use as contributors to the recession and pocketing. Gentler brushing, a sensitive setting, and improved interdental cleaning were advised. Between thirteen and fourteen there is a healing ulceration consistent with toothbrush trauma. Thirty has recession and Grade I furcation involvement, with heavy right-sided occlusal loading. A periodontist consultation is planned, with possible gingival grafting at thirty. Thirty-one bears heavy posterior pressure; bite adjustment is planned to reduce the force concentration. No periodontal stage or grade was assigned.

Clinical content that must survive

- Generalized: At least 4 mm alveolar bone loss; gingival recession with root-surface exposure.
- Generalized: Areas of 4–5 mm pocketing with limited bleeding and inflammation; scaling and root planing planned.
- Generalized: Inadequate interdental cleaning and forceful electric-toothbrush technique identified as contributing factors; gentler brushing, sensitive setting, and improved interdental cleaning advised.
- #13–#14 interproximal area: Healing ulceration consistent with toothbrush trauma.
- #30: Recession, Grade I furcation involvement, and heavy right-sided occlusal loading; periodontist consultation planned, with possible gingival grafting.
- #31: Heavy posterior occlusal pressure; bite adjustment planned to reduce force concentration.
- No periodontal stage or grade assigned.

What this reveals

“Generalized gum disease; deep cleaning and grafting recommended” drops the measurements, the local exceptions, the causal assessment, and the difference between a planned consultation and possible grafting.

This case is primarily periodontal, with associated occlusal findings. A good final note can distribute the content across examination, assessment, education, and plan sections. The relationships must remain clear after that organization.

Fail examples. Calling the 4 mm bone loss a probing depth; changing limited bleeding to no bleeding; applying the furcation finding to every tooth; calling the graft completed or definite; inventing a periodontal stage.

Cases 3 and 4 Measurements and tooth identity

Case 3 Correct the value and preserve the site

Encounter. At thirty mesio-buccal, probing is six millimeters, recession two millimeters, and bleeding is present. Correction: that probing depth is five, not six. At thirty disto-lingual, probing is four millimeters with no bleeding. At thirty-one mesio-buccal, probing is six millimeters with bleeding and suppuration. The clinician plans periodontal reassessment in six weeks. No stage or grade is assigned today.

Expected clinical content. Retain all three sites separately, including the corrected measurement, the distinct recession value, and each bleeding status.

Site	Probing depth	Other findings
#30MB	5 mm	Recession 2 mm; bleeding present
#30DL	4 mm	No bleeding
#31MB	6 mm	Bleeding and suppuration present

Plan: Periodontal reassessment in six weeks. No stage or grade assigned.

Fail examples. Keeping 6 mm at #30MB, moving suppuration to #30, filling unstated recession values with zero, or adding a calculated attachment level or disease stage.

Why transcription is insufficient. The original and corrected numbers both appear in the conversation. The final record must identify the corrected value and keep each measurement with its proper site and meaning.

Case 4 Primary teeth are not permanent teeth

Encounter. Primary tooth A has mesial-occlusal caries. Primary tooth J has distal-occlusal caries. Permanent tooth three has an existing occlusal composite with no fracture seen. Monitor mesial fourteen. Restorations for A and J are planned for a later visit. No restoration was performed today. The parent agrees to those restorations and asks that they be scheduled after the school trip; no appointment date is set.

Expected clinical content. #AMO caries and #JDO caries; restorations planned for both at a later visit. #30 existing composite, with no fracture seen. #14M monitor. No restoration today. Parent agrees to planned care and requests scheduling after the school trip; date not set.

Fail examples. Turning #A into #3; treating A as a shade; grouping both primary teeth under the same surfaces; calling the composite new; recording a booked appointment or patient consent that the encounter did not establish.

Why transcription is insufficient. The note must preserve two dentitions, two different surface patterns, an existing restoration, and a parent's scheduling request without mixing their roles.

Cases 5 and 6 Judgment and completed treatment

Case 5 A complete SOAP note with uncertainty

Encounter. Patient reports lingering cold pain at twelve for three days and no spontaneous pain. Cold testing at twelve reproduces pain for thirty seconds; percussion is not tender. The clinician's assessment is possible irreversible pulpitis, pending endodontic evaluation before a definitive diagnosis. The patient accepts the referral but defers treatment until the evaluation. No root canal treatment was performed today. Return sooner if pain worsens.

One acceptable SOAP note. The following is illustrative wording, not a vendor output.

- S — Patient reports lingering cold pain at #12 for three days; no spontaneous pain.
- O — #12: Cold testing reproduces pain for 30 seconds; non-tender to percussion.
- A — Possible irreversible pulpitis at #12; definitive diagnosis pending endodontic evaluation.
- P — Endodontic referral accepted. Patient defers treatment pending evaluation. Return sooner if pain worsens. No root canal treatment performed today.

Fail examples. Changing possible pulpitis to confirmed pulpitis; omitting the negative percussion result or treatment deferral; adding an antibiotic, radiographic lesion, or completed root canal.

Why transcription is insufficient. Symptoms, test results, clinical interpretation, and the patient's decision have different roles. A heading labeled Assessment does not authorize the AI to diagnose.

Case 6 Record what was administered and completed

Encounter. Patient reports penicillin allergy with hives. One carpule of two-percent lidocaine was administered for fourteen. A second carpule was requested but not administered. We completed the mesial-occlusal composite on fourteen, shade A2, using rubber-dam isolation. A crown for fifteen was discussed; the patient deferred that decision because of cost. No antibiotic was prescribed.

Expected clinical content. Patient-reported penicillin allergy with hives. #14MO composite completed, shade A2, with rubber-dam isolation. One carpule of 2% lidocaine administered for #14; a second requested but not administered. Crown for #15 discussed; decision deferred because of cost. No antibiotic prescribed.

Fail examples. Two carpules administered; allergy omitted or replaced with no known allergies; invented epinephrine concentration, route, volume, or total dose; a crown on #15 recorded as accepted or completed.

Why transcription is insufficient. A medication request is not administration. A treatment discussion is not acceptance. The note must retain those distinctions alongside the actual procedure details.

Cases 7 and 8 Implants and orthodontics

Case 7 Device specifications and canceled sedation

Encounter. At implant site twenty, a four-point-three by ten-millimeter implant was placed today. Final insertion torque was thirty newton-centimeters. Correction: thirty-five newton-centimeters was the final torque at twenty. Placement at implant site twenty-nine was discussed for next month only. Two milligrams of midazolam were prepared but not administered; sedation was canceled. A postoperative review for site twenty is planned in one week.

Expected clinical content. Implant placed today at Site 20; dimensions 4.3 × 10 mm; corrected final insertion torque 35 Ncm. Implant placement at Site 29 discussed for next month, not completed. Midazolam 2 mg prepared, not administered; sedation canceled. Postoperative review for Site 20 in one week.

Fail examples. Recording 30 Ncm as final; treating implant dimensions as tooth numbers; transferring Site 20 treatment to Site 29; documenting midazolam administration; inventing a manufacturer, lot number, bone graft, or consent.

Why transcription is insufficient. The numbers describe different things: a site, device dimensions, torque, a prepared medication, and timing. The final note must preserve all of those roles.

Case 8 Aligner design is not a performed procedure

Encounter. Patient reports wearing aligners sixteen hours daily. A scan was completed today for refinement design. The measured space between two and three is zero-point-two millimeters; correction, zero-point-three. Between three and four it is zero-point-one millimeters. These are space measurements for the design. No interproximal enamel reduction was performed. The clinician advised twenty-two hours of daily wear. Refinement delivery is planned after design approval; no delivery date is set.

Expected clinical content. Patient-reported wear: 16 hours/day. Scan completed for refinement design. Corrected #2–#3 space: 0.3 mm; #3–#4 space: 0.1 mm. No interproximal enamel reduction performed. Advised wear: 22 hours/day. Refinement delivery planned after design approval; date not set.

Fail examples. Stating the patient already wears aligners 22 hours/day; retaining the superseded 0.2 mm; summing the pair measurements and losing their locations; recording IPR or refinement delivery as completed.

Why transcription is insufficient. The note must separate current behavior from advice, design measurements from treatment, and a completed scan from a future delivery.

Cases 9 and 10 Uncertainty and context

Case 9 A lesion with a conditional next step

Encounter. Patient reports a sore on the right side of the tongue for ten days and says a night guard rubs there. Examination shows a six-by-four-millimeter ulcer on the right lateral tongue, with no palpable induration. The clinician considers local trauma a possible cause but has not confirmed the cause. Review in two weeks; biopsy will be considered if it persists. No biopsy was performed today.

Expected clinical content. Patient-reported tongue sore for ten days and night-guard rubbing. Right lateral tongue: 6 × 4 mm ulcer; no palpable induration. Possible local trauma, cause unconfirmed. Review in two weeks; biopsy considered if persistent. No biopsy performed today.

Fail examples. Changing the location to the cheek or left tongue; labeling the lesion definitely traumatic or benign; adding a cancer diagnosis; replacing the conditional biopsy plan with a completed procedure.

Why transcription is insufficient. The patient's explanation and the clinician's tentative assessment are separate. The note must preserve the negative finding and the condition that triggers the next decision.

Case 10 Preserve a conflict without inventing a resolution

Encounter. Current patient: an upper-left posterior tooth has a chipped crown; its number is not identified. The clinician plans reassessment before deciding on replacement. Two entries for this same visit disagree: one says the distal margin of twelve is intact, and the other says it is open. There is no correction, timestamp, or other information to resolve the disagreement. The current visit then ends. A teaching discussion about a different fictional patient describes an abscess at thirty. The current patient's scheduling field says root canal thirty; it is an administrative entry, not evidence of findings or care today.

Expected clinical content. Upper-left posterior tooth: chipped existing crown; specific tooth unresolved; reassessment before a replacement decision. #12D margin: conflicting intact/open entries, requiring clinician clarification. No #30 abscess or completed root canal attributed to this patient.

Fail examples. Assigning the chipped crown to #12 or #14; selecting one margin finding as confirmed; hiding the conflict; importing the teaching case or scheduling entry as clinical evidence.

Why transcription is insufficient. A usable record preserves what is known, makes a material conflict visible, and keeps unrelated clinical speech outside the patient's note. A review warning must accompany the delivered note to count.

Use the results in your practice

Keep the public-suite score separate from tests on your own authorized cases. Include routine and difficult encounters, record every first result, and measure review and correction time. These ten cases test documentation quality; they do not certify clinical safety, privacy, integration, or specialty-wide performance. Dentist review remains necessary.

Published by DentScribe.AI. The same criteria apply to every product. Expected-content examples are not measured product results. No proprietary implementation details are required or disclosed.

F1 A consultation that keeps returning to the teeth

Source-based challenge. Selected excerpts appear in their original order from one complete consultation transcript. Identifying and unrelated passages are omitted. Bracketed omissions are editorial; no timestamps or speaker identities are invented. This is an excerpt set, not the full transcript.

Submit only the numbered excerpts below, together as one input. The next page is the reviewer's answer guide. Use the same pass rules as Cases 1–10.

1 Early patient concern "I know that I need a cavity fill in number 14. Okay. Back side number 14."

2 Later patient decision "And I don't want any implants. Okay, fine. I'm not going through that. So then implants off the table."

3 Examination discussion "Now, this is tooth number 14. Yeah. This is the cavity. Yeah. We'll kind of talk through that a little bit. It started on your root surface right here."

4 After discussion of other teeth "Well my mouth is super dry right now because I've been wearing a mask so I'm just saying this is not how it feels normally."

5 Wear assessment "Where you've got these scooped out areas where the enamel sits higher than the dentin underneath it, that is acid."

6 Later rapid charting "Okay, 14, distal." [Intervening passage omitted.] "And then 19 and 20, distal, decayed. 19 has a PFM."

7 Return to the original tooth "Now, this is the backside of 14, so this is where your cavity is. If I can have access, meaning if my handle fits, I might be able to fill this from the side and stay away from the top."

8 Near the end of the treatment discussion "So I really need three fillings? Yes, but I'd also replace this crown. Okay, and because it's metal? Because it has the cavity underneath it."

9 After appointment discussion "Also fill, why is it not keeping me selected? I would also fill this area here, which is simple." [Intervening passage omitted.] "I wouldn't get you numb, and I wouldn't drill anything else away because there's enough that's already gone."

10 Qualification that follows "Unless I put air on there and you're like, that's sensitive, then I would."

Unrelated conversation and other clinical passages occur between these excerpts. For this public test, score only the clinical evidence supplied above; do not borrow missing details from an assumed full record.

F1 What the assembled note must preserve

The case tests whether a later fragment completes an earlier finding. A note written from the opening concern alone will miss the broader examination and decisions. A note written from the closing treatment discussion alone will lose much of the anatomy.

One acceptable organization

Subjective. Patient presents seeking a restoration on the back of #14. Reports current dry mouth attributed to mask use and says this is not the usual sensation. Patient declines implants.

Objective. #14D: Root-surface caries. #19D and #20D: Caries. #19: Existing PFM crown. Scooped wear areas with enamel higher than underlying dentin, without specific tooth numbers in the supplied excerpt.

Assessment. Clinician attributes the described wear pattern to acid. Decay beneath an existing crown is given as the reason for recommending replacement; metal alone is not the reason. The supplied excerpts do not conclusively identify which crown is referenced in that replacement discussion.

Plan. Restoration of #14, with side access considered if access permits. Three fillings and crown replacement discussed, followed later by an additional area proposed for restoration. The excerpt does not identify that additional site or explicitly reconcile the final procedure count. Additional-area restoration discussed without further drilling and without anesthesia unless air testing produces sensitivity. These are proposed procedures, not documented completed care.

How scattered evidence becomes useful

- Excerpts 1 + 3 + 6 + 7: A patient's initial request becomes a tooth-specific finding and conditional restorative plan.
- Excerpts 4 + 5: A patient-reported sensation and the clinician's wear assessment remain distinct; neither establishes chronic dry mouth or a reflux diagnosis.
- Excerpts 6 + 8: Keep the numbered decay findings and crown history, but do not guess the replacement target from proximity alone.
- Excerpts 8 + 9 + 10: Preserve the late addition and anesthesia condition; do not stop at the earlier three-fillings discussion.

Pass requires. All of the above clinical meaning, including explicit uncertainty, must survive. Clear equivalent organization is acceptable. Do not invent tooth numbers, a final total of four completed fillings, or anesthesia administration.

Why this matters. The dentist never dictates a finished SOAP report. The scribe must connect remarks across the encounter while recognizing what the available words cannot establish.

F2 Extractions mixed with unrelated clinical stories

Source-based challenge. These de-identified excerpts retain the order of a complete pediatric extraction transcript. Long stretches of small talk and identifying details are omitted. The transcript has no reliable speaker labels or timestamps; none are supplied as clinical facts.

Submit the numbered excerpts together as one input. The gaps are deliberate: the scribe must keep track of the current procedure when the conversation wanders.

1 Explanation before care “So what you’re going to feel is pressure. So you’ll feel me kind of pushing, but it should never hurt. So each time I do something, I’ll ask you to make sure that it’s just pressure.”

2 Numbing begins “So let’s start on the left. So what I’m going to do is we’re going to put some numbing jelly up here.”

3 After unrelated conversation “All right, bud. So, what you’re going to feel is just kind of a little pinch up here. I’m going to shake your cheek just like this, okay? Just a tiny pinch up there.”

4 Later response during care “And that doesn’t hurt, right? No.” [Intervening passage omitted.] “Okay, but it’s not painful, right? It just feels like pressure? Yeah.”

5 During family small talk “Okay, one down. Good job, [name removed]. All right, we’re going to do the other one.”

6 A different patient enters the conversation “Now, I had a patient earlier today.” [Unrelated clinical story omitted.] “It’s really uncommon that I see super localized abscesses that can be drained. This was one of them. And it was gushing, and it was wonderful.”

7 Return to the current procedure “We’re doing the other side now. So, this one. Hey, that one’s looser. Remember? That one’s looser. You said we were going to do two. I was wrong. We have four. Do you want to do two at a time? No. I thought it was just these two. We’re here. Let’s just do it.”

8 Much later “You have four teeth pulled out.”

9 Postoperative instructions “So you take one square from either side, take a square, you or you, fold it in half, and then roll it up sort of like a beach towel, okay?” [Intervening passage omitted.] “Now since it’s baby teeth, they caulk pretty fast.” [Intervening passage omitted.] “Bite them.”

10 Closing advice “Softer foods today, if you soar, ibuprofen, um, and that’s it.”

Speech-recognition wording such as “caulk” and “soar” is retained to show the source’s noise. The benchmark checks supported clinical meaning; it does not require copying transcription errors into the note.

F2 What the assembled note must preserve

One acceptable organization

Subjective and response. During the extraction procedure, the patient reports pressure without pain when asked.

Observed status. A tooth on the later-treated side is described as looser. No exact primary tooth identifiers are supplied.

Treatment performed. Four primary teeth extracted, based on the later completed-procedure statement. Numbing jelly and a local numbing injection are described; agent, concentration, dose, and exact tooth sites are not established in the excerpt set.

Education and after-care. Pressure versus pain explained and checked during care. Gauze folding, placement on both sides, and biting instructions provided. Softer foods advised for the day; ibuprofen advised if sore, without a stated dose.

How scattered evidence becomes useful

- Excerpts 1 + 4: An explanation before treatment and the patient's response during treatment belong together, despite the intervening conversation.
- Excerpts 2 + 3: Record the supported numbing steps without inventing a standard anesthetic or dosage.
- Excerpts 5 + 7 + 8: "One down," the change from two to four, and the later completion statement establish the actual procedure count.
- Excerpt 6: The abscess story explicitly concerns another patient. It must not enter this child's diagnosis or treatment.
- Excerpts 9 + 10: The closing remarks supply the after-care plan, long after the initial procedure explanation.

Pass requires. Four completed primary-tooth extractions, supported numbing and patient-response details, and the supplied after-care instructions. Do not document abscess drainage for this child, assign unstated tooth letters, or invent medication details or formal consent.

Complete the scattered-conversation result

F1 result _____ Reason or correction _____

F2 result _____ Reason or correction _____

Scattered-conversation passes ____ / 2

Clinical errors ____ Organization only ____ No usable note ____

A focused test shows whether a scribe can preserve difficult details. A scattered-conversation test shows whether it can assemble those details across an encounter. Report both. For your practice trial, also use complete authorized transcripts without pre-sorting or trimming the clinical conversation; report that trial separately from this public excerpt suite.

Source note: F1 and F2 were selected after reviewing complete consultation and extraction transcripts supplied by DentScribe. Public excerpts omit identifying material and preserve the selected passages' order. They are not complete patient records, audio-verified transcripts, or measured outputs from any vendor.