The Documentation Minute: A Multi-Site Analysis of Time Spent on Clinical Documentation in Psychiatry, and the Effect of AI-Assisted Notetaking

Lotte Kjær Svalberg

Cand.psych.; stud.med.; Clinical Product Manager, Aisel

Abstract

Background. Clinical documentation is a major driver of clinician burnout and a structural limit on how many patients a psychiatric service can see. The published evidence on AI-assisted documentation has grown quickly, but it is concentrated in US primary care, focuses almost entirely on physicians, and evaluates a single task type: the post-consultation note. Longer, more specialised document types remain largely unstudied.

Methods. We analysed 202 time documentation entries from eight clinic and practice settings in Denmark and the United Kingdom, spanning public and private psychiatric services and 16 documented task types (clinical notes, discharge summaries, medico-legal reports, benefits reports, court reports, and others). Entries were recorded under three conditions: no AI assistance, AI assistance using Aisel, and AI assistance using other tools. Time was captured by stopwatch or retrospective estimate.

Results. Clinical notes were the only document type with an adequate sample on both sides of the comparison: 14.5 minutes without AI assistance (n=43) versus 5.5 minutes with Aisel (n=37), a 62% reduction in mean time; medians were 10.0 and 5.0 minutes, a 50% reduction (Mann-Whitney U=1377, p<.001, rank-biserial r=0.73). This is consistent in direction and size with an independent proof-of-concept study in an NHS psychiatric clinic, in which AI-assisted documentation took 45% of manual documentation time. Descriptive data across all 16 task types (n=152 non-AI entries) show that documentation burden is concentrated in long-form outputs: Mental Health Act reports (57 minutes), medico-legal reports (50 minutes), and court reports (46 minutes). None of these has yet been evaluated with AI assistance in the published literature.

Conclusion. A substantial, cross-site reduction in documentation time for psychiatric clinical notes is consistent with the small body of psychiatry-specific evidence. The largest unexplored opportunity, however, lies in the document types AI has not yet been studied on. We discuss implications for clinical scalability, employee well-being, and service economics, and outline the methodological limitations that should guide larger evaluations.

Keywords: AI-assisted documentation; psychiatry; clinical scalability; documentation burden; ambient AI scribe

1. Introduction

Clinical documentation takes up a large share of the working day. Across specialties, clinicians spend an estimated two hours on electronic health record (EHR) tasks for every hour of direct patient contact, with much of that work displaced into personal time, so-called "pajama time" (Saag et al., 2019; Barr et al., 2024). Documentation burden is among the strongest predictors of clinician burnout, and interventions that reduce it have been associated with lower stress, better usability ratings, and in some cases reduced staff turnover (Alobayli et al., 2023). In psychiatry, where the clinical work itself is language-based (history-taking, risk assessment, therapeutic dialogue), documentation competes directly with patient care for the same cognitive and temporal resources: time spent writing about one session is time not available for the next.

The burden is not evenly distributed across medicine. In a nationally representative survey of US physicians, psychiatrists reported spending 10.6 hours per week (20.3% of working time) on administrative tasks, the highest proportion of any specialty surveyed (Woolhandler & Himmelstein, 2014). This fits the nature of the work: psychiatric assessments and formulations require narrative depth, and the documentation that follows a session is rarely a brief, structured note. A related question, which our dataset cannot answer because it measures time rather than interaction quality, is what happens to the consultation itself when the clinician types less during it. An emerging literature has begun to address this, and we return to it in section 4.4.

Large language model (LLM) tools that draft clinical notes from consultation audio, commonly termed ambient or AI scribes, have produced a fast-growing evidence base. A 2025 systematic review and meta-analysis pooling 23 studies found a moderate, statistically significant reduction in both documentation burden and documentation time with AI tool use (standardised mean difference ≈ −0.71 to −0.72) (Zhao et al., 2025). AI scribe adoption has also been linked to improvements in physician-reported burnout and usability (Shah et al., 2025) and, in one large single-site study, to increased billed clinical activity (Holmgren et al., 2026).

Three features of this evidence base limit how confidently it generalises to psychiatric practice. It is concentrated in the United States and in general medical or primary care settings; the most recent systematic review included only one psychiatric study, which used simulated cases with a sample of eight (Zhao et al., 2025). It is almost exclusively physician-centred, leaving psychologists, nurses, and other professions who carry much of the documentation work unmeasured. And it focuses on a single task type, the note written immediately after a consultation, while the longer documents psychiatric clinicians routinely produce (discharge summaries, medico-legal reports, disability and benefits assessments, tribunal and court reports) have essentially no published evaluation of AI assistance at all.

One exception offers a useful point of comparison. A proof-of-concept study in an NHS child and adolescent mental health service found that AI-assisted documentation reduced median time per encounter from 27 to 10 minutes, with AI-assisted notes taking on average 45% of manual documentation time, alongside high patient acceptance (Stanton et al., 2025). As we report below, our own data, collected independently in different clinics and countries and with a different AI tool, show a reduction of similar size for the same underlying task. We return to this convergence in the discussion.

This paper has two aims. The first is descriptive: to map, across 202 documentation time entries from eight settings, two countries, and 16 task types, where documentation time is spent in psychiatric practice. The second is evaluative but deliberately narrow: to report, with the necessary cautions about sample size and cross-site comparison, the one comparison in our data (clinical notes) where AI-assisted and unassisted documentation can be measured against each other. We close by discussing what the findings mean for four outcomes: clinical scalability, employee satisfaction, profitability, and patient experience.

A note on provenance: this analysis was conducted by a member of the team behind Aisel, and all Aisel-assisted entries come from a single clinic using the tool. We have tried to write it to be checkable rather than persuasive: the measurement methods and confounders are described in full, the limitations are collected in section 6, and we flag at each point whether a claim rests on our own data or on published work. Readers should weigh the findings with our interest in them in mind.

2. Methods

Data collection. Entries were contributed by clinicians and administrative staff across eight clinic and practice settings in Denmark and the United Kingdom, spanning public psychiatric services and private practices. For each documentation task, contributors recorded the clinic, task type, time spent (minutes), measurement method, AI-use condition, staff role, and an optional free-text comment. Clinic and task type were selected from a maintained list; 16 task types appear in this dataset.

Measurement methods. Time was captured in one of two ways: real-time stopwatch timing or retrospective estimation. 42 of 202 entries (21%) are stopwatch-timed, all contributed by a single clinic; the remaining 160 (79%) are retrospective estimates, contributed mainly by NHS and private-practice clinicians in the UK and Denmark. This distinction matters for how the results should be read, and we return to it in the Limitations. One point about how the retrospective estimates were collected: when a clinician reported a range rather than a single figure (for example, 10 to 20 minutes depending on complexity), the range is retained as a single observation with a recorded minimum and maximum bound. 52 of 202 entries (26%) were reported this way. Analyses below use the range midpoint as the point estimate; the minimum- and maximum-bound averages are reported alongside key figures as a sensitivity check on how much a given average could shift depending on where within a reported range the true value fell.

Analysis. Throughout, we report descriptive statistics (means and sample sizes) by task type, clinic, AI-use condition, and staff role; for the clinical notes comparison, where the distributions are notably skewed, we also report medians and interquartile ranges. For that comparison, the only one with an adequate sample on both sides, we additionally report a Mann-Whitney U test with a rank-biserial correlation, and a Welch's t-test with Cohen's d and a 95% confidence interval for the mean difference, a supplementary check on whether these two samples differ by more than sampling variation, and by how much. Given the observational, non-randomised, multi-site design and the confounding of AI-use condition with clinic, profession, and measurement method (see Limitations), none of this establishes a causal effect: it shows that the samples reliably differ and by how much, not why. The design, not the

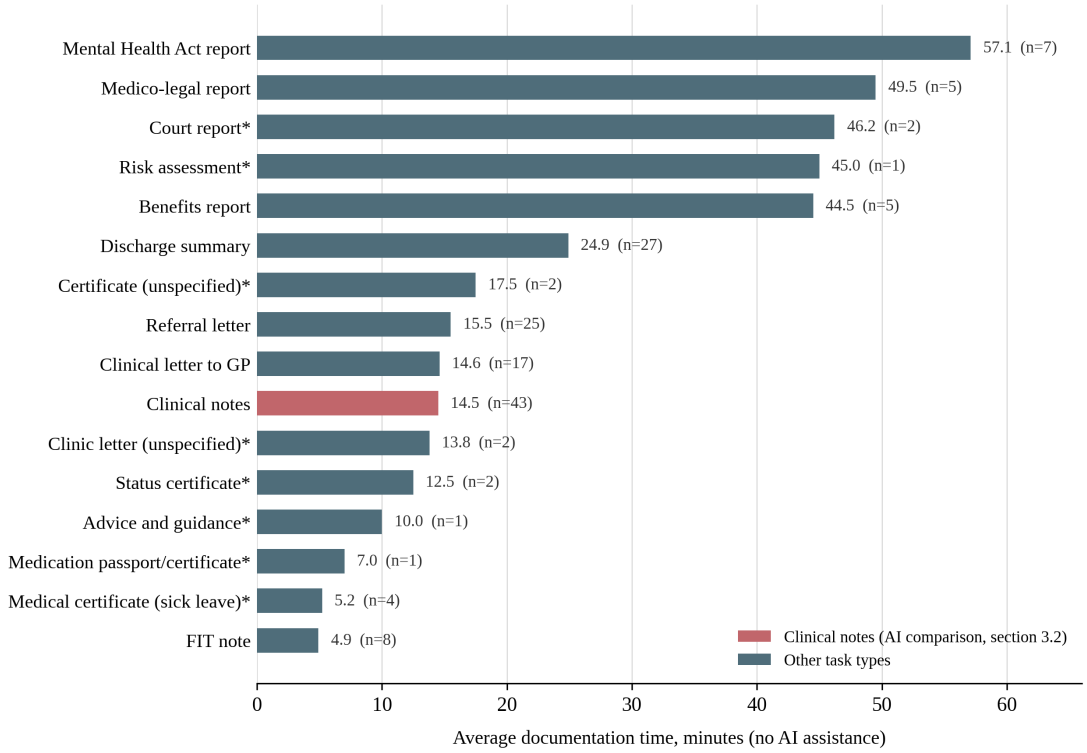
choice of statistic, is the binding constraint here. All other comparisons remain purely descriptive, without significance tests.

3. Results

3.1 The documentation burden landscape

Across the 152 entries recorded without AI assistance, average documentation time varies more than tenfold by task type (Figure 1). The longest times are concentrated in long-form, legally or administratively consequential outputs: Mental Health Act reports, medico-legal reports, and court reports each average 46 to 57 minutes. The shortest are brief, templated outputs such as sick-leave certificates and fit notes, at around 5 minutes. Clinical notes, the task type at the centre of the AI comparison below, sit near the middle of this range at 14.5 minutes.

Several of the longest-duration categories (risk assessment, advice and guidance, status certificate, court report, clinic letter (unspecified)) have samples too small to treat as stable estimates and are included for completeness rather than as reliable benchmarks. The remaining high-duration categories, medico-legal reports (n=5), Mental Health Act reports (n=7), and benefits reports (n=5), are discussed below as a distinct opportunity area; their samples are modest and the averages should be read as indicative rather than precise, but none of them has been evaluated with AI assistance in the literature we reviewed.



** Sample too small (n<5) to treat as a stable estimate; shown for completeness.*

Figure 1. Average documentation time by task type, no AI assistance (n=152). Clinical notes, the task type compared with and without AI in section 3.2, is highlighted. Categories marked with an asterisk have samples too small (n<5) to treat as stable estimates.

3.2 The clinical notes comparison

Clinical notes are the only task type in this dataset with a workable sample on both sides of an AI comparison. Without AI assistance, clinical notes averaged 14.5 minutes (n=43, median 10.0, IQR 7.5–20.0; averaging clinicians' low-end estimates gives 13.0 minutes, averaging their high-end estimates gives 15.9 minutes); with Aisel, 5.5 minutes (n=37, median 5.0, IQR 4.0–7.0), a 62% reduction (57.9%–65.5% across that low-to-high sensitivity range). Given the right-skewed distribution, particularly on the non-AI side, the median and IQR describe a typical single note more faithfully than the mean: on medians, the reduction is 50% (10.0 to 5.0 minutes). The mean answers a different question. Total documentation time across a caseload is the number of documents multiplied by the mean, not the median, so the 62% reduction in mean time is the figure that bears on service capacity, while the 50% median reduction is closer to what a clinician would expect to feel on a typical note. We report both. A Mann-Whitney U test on these two samples returns U=1377, p<.001, rank-biserial r=0.73 (a large effect: a randomly chosen non-AI entry exceeds a randomly chosen Aisel entry about 85% of the time, versus 12% the reverse); a Welch's t-test returns t(46.9) = 5.51, p<.001, mean difference 9.0 minutes (95% CI 5.7–12.3), Cohen's d=1.15, also a large effect by conventional benchmarks. As throughout, these describe how reliably and how much the two samples differ, not that AI assistance caused the difference independent of the confounders discussed below. This is close in direction and size to the only comparable published figure: the NHS child and adolescent psychiatry proof-of-concept in which AI-assisted documentation took 45% of manual documentation time (Stanton et al., 2025). Two independent measurements, made in different countries, health systems, and tools, arriving at a similar-sized reduction for the same task is worth noting, though the finding remains preliminary.

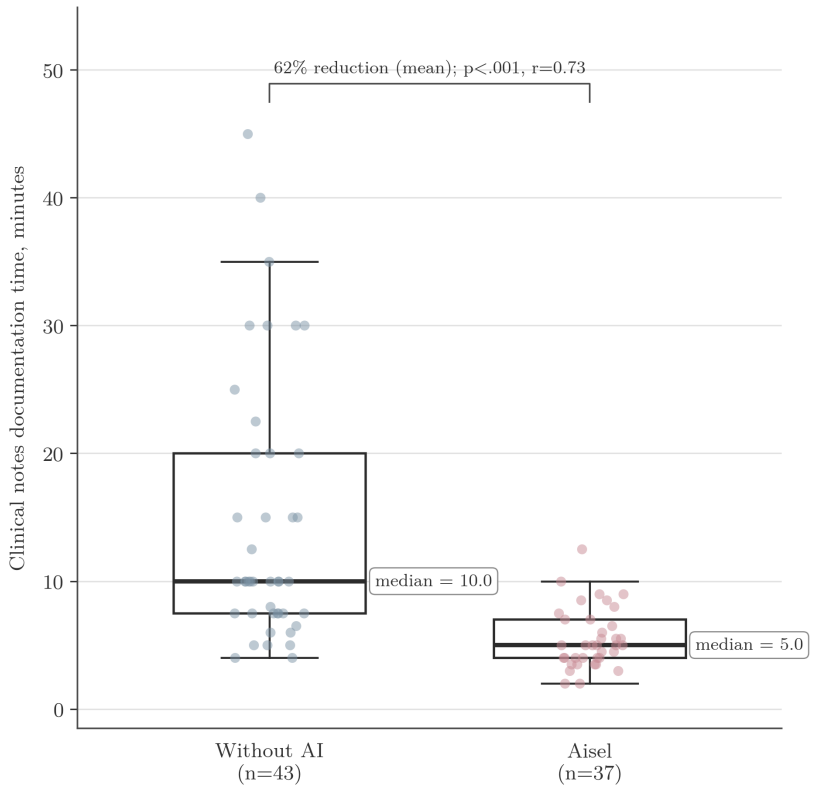


Figure 2. Clinical notes documentation time, without AI (n=43) versus with Aisel (n=37). Box shows median and interquartile range; points are individual entries, jittered horizontally for visibility.

An important caution applies, flagged in the Methods: 39 of the 43 no-AI clinical notes entries are retrospective estimates, while all 37 Aisel entries are stopwatch-timed and from a single clinic. Retrospective time estimates are a well-documented source of bias in time-use research and are not directly comparable with stopwatch timing on a minute-for-minute basis. The statistical test above confirms the samples differ reliably, but it cannot separate that difference from the confound between AI-use condition, clinic, and measurement method. The 62% figure should therefore be read as a cross-site signal consistent with the wider literature, not as a controlled measurement of AI's causal effect on documentation time. A secondary cut points the same way without resolving this caution. Restricting the comparison to psychologists, the profession represented in both conditions, documentation time was 8.2 minutes without AI (n=7) versus 5.5 minutes with Aisel (n=37). At n=7, this subsample is too small to lean on: it is offered only as a weak, indicative consistency check, not as corroborating evidence in its own right. Because nearly all Aisel entries were logged by psychologists, it also re-slices the same underlying data rather than confirming it independently.

Thirteen entries used AI tools other than Aisel for clinical notes and other task types, averaging around 3.7 minutes. The sample is too small and too heterogeneous in tooling to interpret and is noted for completeness only.

4. Discussion

We organise this discussion around the four practical outcomes that motivated the analysis: clinical scalability, employee satisfaction, profitability, and patient experience. The evidence behind each differs. Some rest on data collected in this study or on close published analogues; others rest on mechanisms reported elsewhere in the literature. We note which is which in each case.

4.1 Clinical scalability

Figure 1 is itself a scalability argument. In a service whose mix of work resembles the one captured here, medico-legal reports, Mental Health Act reports, and benefits reports can each consume 30 to 60 minutes or more, on top of clinical notes, referral letters, and discharge summaries. Every minute spent producing a report is a minute not spent seeing the next patient, and in systems already reporting long waits for psychiatric assessment, the ceiling on capacity is set as much by documentation as by consulting rooms. One contributing clinic remarked, in an internal comment collected alongside the dataset (anecdotal, and not independently verified), that “direct patient contact only takes up 40% of time, which currently makes scaling clinics difficult.” We report this as anecdote rather than data, but it is consistent with the shape of Figure 1.

The clinical notes finding is where we can be most concrete: a 62% reduction, if it generalised to even a fraction of the longer document types above, would carry considerably larger capacity implications than the notes figure alone suggests. This is a hypothesis for future study, not a finding of this one, since AI assistance in our dataset was only ever used for clinical notes.

4.2 Employee satisfaction

Psychiatry's administrative burden is the largest of any physician specialty measured, at roughly a fifth of total working time (Woolhandler & Himmelstein, 2014), and documentation burden is one of the most consistent predictors of burnout in the wider literature (Zhao et al., 2025). Within our own data, the one same-profession comparison available points the same way: psychologists' clinical notes time fell from 8.2 to 5.5 minutes with Aisel (n=7). This is consistent with published evidence linking reduced documentation burden to better usability ratings and lower burnout scores, even where the objective time savings are modest (Shah et al., 2025). A simulation study of psychiatric consultations found a drop in perceived workload, alongside higher self-rated performance and lower frustration, when clinicians used an AI scribe rather than typing in real time (Nawaz et al., 2025, preprint). The mechanism, less time on documentation and less burnout, is established in the literature, and our data are directionally consistent with it, though this study was not designed to measure burnout or satisfaction directly.

4.3 Profitability

This outcome requires the most care in translation, because most of the published economic evidence comes from a healthcare system unlike those that contributed most of our data. In US fee-for-service practice, AI scribe adoption has been associated with measurably higher billed activity, on the order of an additional RVU per physician per week (Holmgren et al., 2026), a mechanism that does not map onto NHS or Danish public psychiatric services, which are not paid per encounter in the same way. A more transferable mechanism is cost avoidance rather than revenue generation. Burnout-driven turnover has been estimated to cost health systems $500,000 to over $1,000,000 per departing physician in replacement and lost-productivity costs, with burnout-attributable costs of roughly $7,600 per physician per year at the organisational level (Han et al., 2019). If reduced documentation burden contributes even modestly to retaining staff, the cost-avoidance argument holds regardless of how a system bills for care. We have not measured this directly, and the dollar figures are US-specific and should not be transplanted uncritically into a UK or Danish cost structure.

4.4 Patient experience

Our data speak to this outcome least directly: we collected no patient-reported outcomes, and nothing in the dataset measures how patients experienced their consultations. What we can offer is a chain of findings from elsewhere in the literature. Attentional presence and eye contact are established correlates of perceived empathy and therapeutic alliance (MacDonald, 2009); in the simulation study, reduced note-taking during consultations was associated with standardised patients rating clinicians as more attentive, less rushed, and more compassionate (Nawaz et al., 2025, preprint); and the NHS proof-of-concept reported high patient acceptance of AI-assisted documentation, with the large majority of patients unbothered by its use (Stanton et al., 2025). This is a literature-supported hypothesis worth testing directly in future work, not a conclusion this study can draw.

5. Where the Opportunity Is Largest

Everything reported so far about AI assistance concerns one task type: clinical notes. Medico-legal reports (49.5 minutes, n=5), Mental Health Act reports (57.1 minutes, n=7), benefits reports (44.5 minutes, n=5), and discharge summaries (24.9 minutes, n=27, the largest sample after clinical notes) together account for a large share of total documentation time in this dataset, and none of them has been tested with AI assistance here or, as far as we found, anywhere in the published literature. In the systematic review discussed earlier, 20 of 23 included studies evaluated only the post-consultation note; discharge summaries appeared in two; the long form report types that dominate the top of Figure 1 appeared in none (Zhao et al., 2025).

This gap is not simply an oversight. These report types differ from clinical notes in ways that could make them either more or less suited to AI assistance. Several follow strongly templated, statutory, or pro-forma structures (Mental Health Act reports and benefits assessments in particular are built around fixed headings and defined information requirements), which language models tend to handle well. They also demand more synthesis across a longer clinical history, more explicit legal or clinical reasoning, and carry higher stakes if summarised inaccurately, all of which raise the bar for acceptable AI assistance. Which of these considerations dominates is unknown, because it has not been studied.

As an illustration, if a reduction of the size observed for clinical notes (62%) applied to medico-legal reports, roughly 31 minutes would be returned per report, on a task currently averaging around 50 minutes. This sizes the gap; it is not a projection, since no AI-assisted data exist for these task types in our dataset or, to our knowledge, in the literature. The narrower conclusion stands on its own: of everywhere documentation time currently goes in psychiatric practice, the largest untapped concentrations sit outside the one task type the evidence base, including our own data, has so far examined.

6. Limitations

This analysis has several limitations, most of which have been flagged as they arose; we collect them here. The most consequential is the study's design, not its statistics: the confounding described next, between AI-use condition, clinic, profession, and measurement method, cannot be resolved by any choice of significance test, effect size, or confidence interval. Those statistics describe how reliably and how strongly the samples we have differ; they cannot supply the randomised comparison we do not have.

The core AI comparison is observational and cross-site, not a controlled trial. Aisel-assisted entries come entirely from a single clinic; the non-AI comparison group is drawn mostly from different clinics, countries, and health systems doing the same nominal task type. Site, patient population, documentation standards, and task complexity are all potential confounders we cannot separate from the effect of AI assistance itself.

Measurement method is confounded with AI-use condition in the same direction. All 37 Aisel entries are stopwatch-timed; 79% of the full dataset, and 39 of the 43 non-AI clinical notes entries specifically, are retrospective estimates. Retrospective estimation is a recognised source of measurement bias in time-use research, and some portion of the observed difference may

reflect method rather than AI assistance. The Mann-Whitney/Welch's t results reported in 3.2 confirm the two samples differ reliably; they do not adjust for this confound.

Where clinicians described a range rather than a single figure (52 of 202 entries, 26%), the observation is retained as a single data point with a recorded minimum and maximum bound. The underlying imprecision of a retrospective range estimate remains, however, so where an average rests substantially on range midpoints we also report the low- and high-bound averages (Figure 1's underlying data and the clinical notes comparison in 3.2) so readers can see how far the point estimate could shift, especially where n is small.

The “other AI” category (n=13) spans multiple unspecified tools and task types and is too small and heterogeneous to support interpretation; it is reported for completeness only.

The dataset has no recorded entry dates, so we cannot assess whether documentation times changed over the collection period, whether AI-assisted times improved with clinician familiarity, or whether seasonal or caseload effects are present.

The sample is a convenience sample of contributing clinics and clinicians, weighted toward one health system (97 of 202 entries, 48%, are from NHS (UK) services), and should not be assumed to represent psychiatric documentation practice more broadly. As noted in the introduction, this analysis was conducted by a member of the team behind Aisel. We have flagged the points where that relationship could bias interpretation.

7. Conclusion

This paper set out to map where documentation time is spent in psychiatric practice, and to report, with appropriate caution, the one comparison in our data between AI-assisted and unassisted documentation. On the first, the picture is clear: documentation burden in psychiatry is large, uneven, and concentrated in long-form outputs (medico-legal reports, Mental Health Act reports, benefits reports, and court reports) that have received almost no attention in the AI documentation literature. On the second, clinical notes show a substantial cross-site reduction in time with Aisel (62%), consistent in direction and rough size with the one independent psychiatry-specific study we found (Stanton et al., 2025), though our measurement carries real limitations around site, measurement method, and sample independence.

Against the four outcomes that motivated this work, the strongest claims concern staff well-being and the size of the documentation burden itself; profitability and patient experience are plausible but rest more on published mechanisms than on anything measured here. The most useful next step is not a broader claim about the clinical notes finding, but a targeted evaluation of AI assistance on the report types that consume the most time and have been studied least: medico-legal, Mental Health Act, and benefits reports.

References

Alobayli, F., O'Connor, S., Holloway, A., & Cresswell, K. (2023). Electronic Health Record Stress and Burnout Among Clinicians in Hospital Settings: A Systematic Review. *DIGITAL HEALTH*.

Barr, W., Peterson, L., & Fleischer, S. (2024). Pajama Time: The Association of EHR Documentation Time with Family Medicine Resident Outcomes. *Annals of Family Medicine*, 22(Supplement 1), 6628.

Han, S., Shanafelt, T. D., Sinsky, C. A., et al. (2019). Estimating the Attributable Cost of Physician Burnout in the United States. *Annals of Internal Medicine*, 170(11), 784–790.

Holmgren, A. J., Fenton, C. L., Thombley, R., et al. (2026). Ambient Artificial Intelligence Scribes and Physician Financial Productivity. *JAMA Network Open*, 9(1), e2553233.

MacDonald, K. (2009). Patient-clinician eye contact: Social neuroscience and art of clinical engagement. *Postgraduate Medicine*, 121(4), 136–144.

Nawaz, F. A., Bokhari, S. A., Usman, F. M., Arshad, Z., Sudhir, M., Krage, R., & Kashyap, R. (2025). Evaluating an Ambient Artificial Intelligence Scribe for Documentation Quality and Efficiency in Psychiatric Consultations: A Simulation-based Study. *medRxiv* [preprint, not yet peer-reviewed]. https://doi.org/10.1101/2025.09.21.25336260

Saag, H. S., Shah, K., Jones, S. A., et al. (2019). Pajama Time: Working After Work in the Electronic Health Record. *Journal of General Internal Medicine*, 34(9), 1695–1696.

Shah, S. J., Devon-Sand, A., Ma, S. P., et al. (2025). Ambient artificial intelligence scribes: physician burnout and perspectives on usability and documentation burden. *Journal of the American Medical Informatics Association*, 32(2), 375–380.

Stanton, N., Aziz, A., Wong, S., Jakhra, S., & Golchinheydari, S. (2025). Evaluating AI Ambient Voice Technology as a Documentation Assistant in Psychiatry – a Proof of Concept Study. *BJPsych Open*, 11(Suppl 1), S6–S7.

Woolhandler, S., & Himmelstein, D. U. (2014). Administrative work consumes one-sixth of U.S. physicians' working hours and lowers their career satisfaction. *International Journal of Health Services*, 44(4), 635–642.

Zhao, J., Liu, H., Chen, Y., & Song, F. (2025). Application of artificial intelligence tools and clinical documentation burden: a systematic review and meta-analysis. *BMC Medical Informatics and Decision Making*, 26, 29.