Journal of Interdisciplinary Engineering, Computing and Automation
Vol. 01, Issue 01, July–December 2026, pp. 1-12

Who Receives the Saving? A Discrete-Event Analysis of Documentation Time Reduction, Clinic Throughput and Schedule Reliability

Dr. Shahida T

Senior Lecturer, Electrical and Communication Engineering, College of Engineering, National University of Science and Technology, Muscat, Oman.

Email: shahida@nu.edu.om

ABSTRACT

Ambient and automated documentation tools reduce the time clinicians spend writing notes. Whether that reduces clinician workload, increases patient throughput, or neither, depends on a scheduling decision rather than on the technology. We quantify the three outcomes and the effect on schedule reliability. We simulated 20,000 clinic sessions of 240 fixed minutes with 12 booked patients, drawing consultation and documentation times from lognormal distributions and splitting documentation between in-visit and after-hours work. Documentation time was scaled by a factor representing the tool, and extra patients were booked into the unchanged session length to represent organisational absorption of the saving. At baseline, clinicians saw 11.17 patients, worked 37.0 minutes after hours and overran the session in 10.2 per cent of cases. Halving documentation time with the schedule unchanged reduced after-hours work to 18.0 minutes, a fall of 51 per cent, and overruns to 1.2 per cent. Booking 4 extra patients returned after-hours time to baseline (40.1 minutes) and raised throughput 33 per cent, but overruns rose from 10.2 to 65.0 per cent and mean overrun from 0.1 to 16.2 minutes. Requiring both after-hours time and overrun rate to remain at or below baseline permitted only 1 extra patient, against the 1 suggested by mean-time arithmetic. The destination of a documentation saving is an organisational choice that the technology cannot make. Variability, not mean time, sets the safe throughput dividend, and it is roughly a quarter of the arithmetic estimate.

Keywords: clinical documentation; ambient scribes; discrete-event simulation; clinician burnout; clinic scheduling; health informatics.

1. INTRODUCTION

Documentation is among the largest identified consumers of clinician time, and a substantial share of it is done outside scheduled hours [1-3,23,26,27]. Tools that reduce it — ambient speech capture, automated note generation, structured templating — are being adopted rapidly, and early evaluations report meaningful reductions in documentation minutes [4-6].

A reduction in documentation minutes is not the same as a reduction in clinician workload, and it is not the same as an increase in capacity. It is a freed resource, and what happens to it is decided by whoever sets the schedule. If the clinic books the same patients into the same session, the saving accrues to the clinician as shorter days. If it books more patients, the saving accrues to the organisation as throughput and the clinician's day is unchanged.

Both are legitimate uses and each has been advocated, but they are mutually exclusive and the choice between them is frequently left implicit [7,8]. A tool is procured on the strength

of a burnout argument [20-22] and deployed into a scheduling system that absorbs the saving automatically, with the result that the burnout outcome is not observed and the tool is judged to have failed.

A second issue is less obvious. Clinic sessions have a fixed length, and consultation times are variable. Adding patients to a fixed session consumes slack, and because overrun depends on the upper tail of the workload distribution rather than its mean, schedule reliability can degrade sharply well before the average day becomes too long [15,18]. A booking level that looks neutral on a time-accounting basis may not be neutral in practice.

We therefore simulated a clinic session and asked where a documentation saving goes, how much throughput it can safely support, and what happens to the reliability of the schedule along the way.

1.1 Contributions

We quantify after-hours time, throughput and schedule overrun under a range of documentation reductions with the schedule held fixed; we determine the number of additional patients that returns after-hours time to baseline; we show that schedule reliability collapses before that point; and we compare the throughput dividend implied by mean-time arithmetic with the dividend that leaves both after-hours time and overrun rate unworsened.

2. METHOD

2.1 Model

Each simulated session is a fixed 240 minutes with 12 patients booked, corresponding to nominal 20-minute slots. Patients fail to attend with probability 7 per cent.

For each attending patient, consultation time and documentation time are drawn from lognormal distributions with medians of 14 and 6.5 minutes. Lognormal forms were chosen because consultation durations are right-skewed, and it is that skew which drives overrun [12,13]. A fraction of 55 per cent of documentation is completed during the visit and the remainder is deferred.

In-session work is the sum of consultations and in-visit documentation. Where it exceeds the scheduled session, the excess is recorded as an overrun and also spills into personal time. After-hours work is deferred documentation plus any spill.

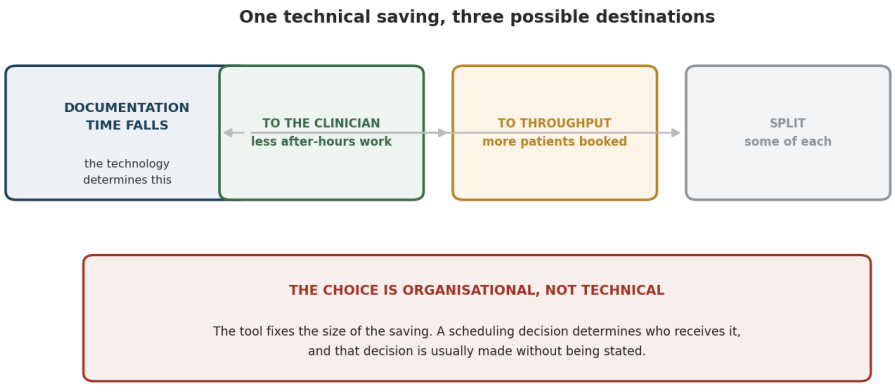


Fig.1: The saving and its three possible destinations.

The documentation tool is represented by a multiplicative factor applied to documentation time only, from 1.0 (no tool) to 0.3 (a 70 per cent reduction). Organisational absorption is represented by booking additional patients into the same fixed session, so that

slots shorten rather than the day lengthening. This is the decision we are studying and it is deliberately the only lever varied. Figure 1 shows the saving and its three possible destinations.

Parameters are given in Table 1. 20,000 sessions were simulated per configuration, using a simulation implemented in Python with NumPy, SciPy and Matplotlib [28-30]. Values are illustrative and represent an outpatient or primary care session rather than a specific service.

Table 1: Simulation parameters.

Parameter	Value	Note
Session length	240 minutes	Fixed; booking more patients shortens slots
Patients booked at baseline	12	Nominal 20-minute slots
Consultation time	Lognormal, median 14 min	Right-skewed; the skew drives overrun
Documentation time	Lognormal, median 6.5 min	Scaled by the tool factor
Documentation done in visit	55%	Remainder deferred to after hours
Non-attendance	7%	Independent per slot
Tool factor	1.0 down to 0.3	Multiplies documentation time only
Sessions simulated	20,000	Per configuration

3. RESULTS

3.1 Baseline and the Effect with the Schedule Unchanged

At baseline the clinician saw 11.17 patients, did 209.4 minutes of in-session work against a 240-minute session, worked 37.0 minutes after hours and overran the session in 10.2 per cent of cases.

Table 2 and Figure 2 reduce documentation time with the schedule held fixed. After-hours work falls almost linearly with the tool factor, from 37.0 minutes to 18.0 at a 50 per cent reduction and 10.8 at 70 per cent. Overruns fall in parallel, from 10.2 per cent to 1.2 and 0.4 per cent, because in-visit documentation is part of in-session work.

Patients seen is unchanged by construction. This is the scenario in which the entire saving reaches the clinician, and it is the scenario on which burnout-reduction business cases are implicitly built.

Table 2: Documentation reduction with the schedule held fixed.

Tool factor	After-hours minutes	Total work minutes	Patients seen	Sessions overrunning
1.00	**37.0**	**245.2**	**11.17**	**10.2%**
0.80	29.1	229.3	11.17	4.7%
0.60	21.7	213.4	11.17	1.9%
0.50	**18.0**	**205.4**	**11.17**	**1.2%**
0.40	14.4	197.5	11.17	0.6%
0.30	10.8	189.5	11.17	0.4%

3.2 Absorbing the Saving Gives It Back

Table 3 and Figure 3 hold documentation at 50 per cent of baseline and book additional patients into the same session. After-hours time rises steadily as slots are added: 18.0 minutes

with no addition, 23.4 with two extra patients, and 40.1 with 4, at which point it has returned to the baseline level of 37.0 minutes.

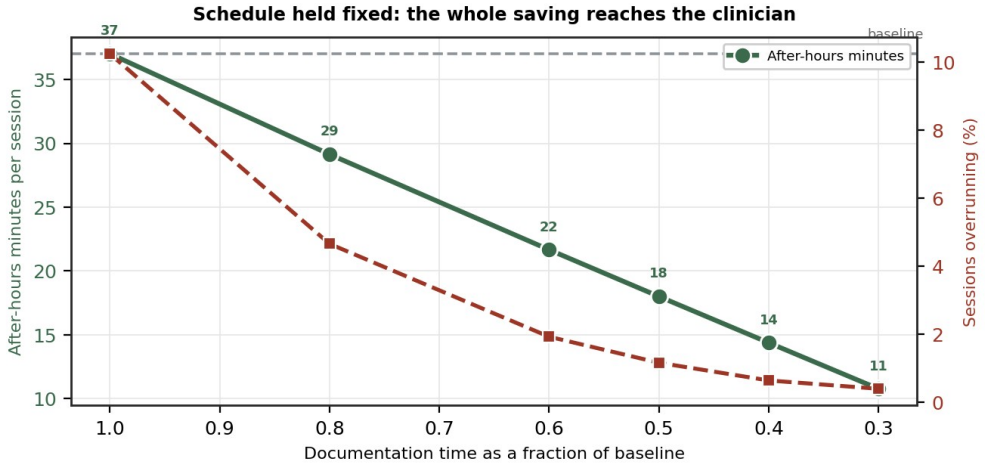


Fig.2: After-hours minutes and overrun rate as documentation time falls, with the schedule unchanged.

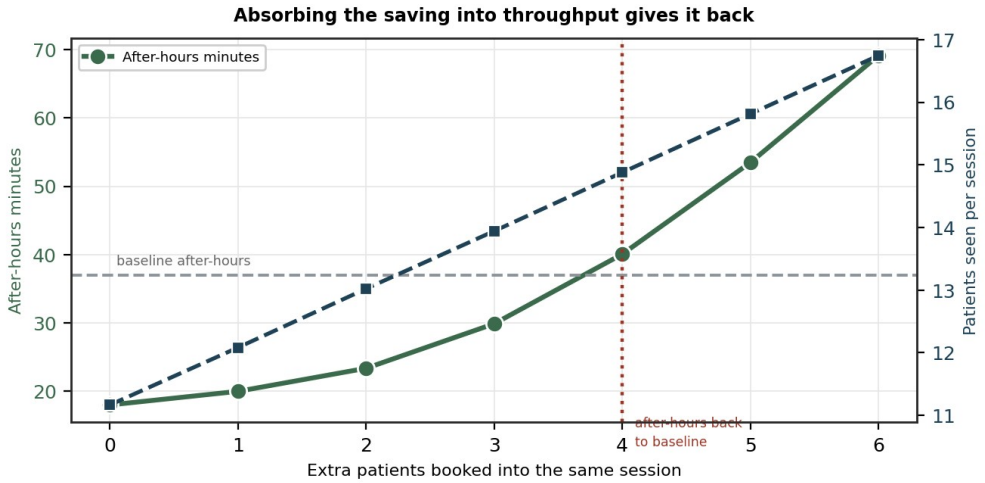


Fig.3: After-hours minutes and patients seen as extra patients are booked into the same session.

At that booking level the clinic sees 14.88 patients against 11.17 at baseline, an increase of 33.2 per cent, and the clinician's after-hours burden is where it started. Total work per session has risen from 245.2 to 273.8 minutes.

This is the arithmetic that a capacity planner would perform, and on its own terms it is correct: a real saving has been converted into real additional access without lengthening the clinician's day. Whether that is the right use of the saving is a question of policy rather than of analysis, and our purpose is not to answer it but to make the trade explicit.

Table 3: Absorption of a 50 per cent documentation reduction into throughput.

Extra patients	Patients seen	After-hours minutes	Change vs baseline	Total work minutes	Sessions overrunning
0	**11.17**	**18.0**	**−51.3%**	**205.4**	**1.2%**
1	12.08	20.0	−46.0%	222.3	5.7%
2	13.02	23.4	−36.9%	239.4	18.8%
3	13.94	29.9	−19.3%	256.4	41.4%
4	**14.88**	**40.1**	**+8.2%**	**273.8**	**65.0%**
5	15.81	53.5	+44.4%	290.9	82.6%
6	16.74	69.1	+86.7%	308.2	93.0%

Baseline for comparison: 11.17 patients, 37.0 minutes after hours, 10.2% of sessions overrunning.

3.3 Schedule Reliability Collapses First

Table 4 and Figure 4 report the overrun rate as slots are added. It rises from 1.2 per cent with no addition to 5.7 per cent with one extra patient, 18.8 per cent with two, 41.4 per cent with three and 65.0 per cent at 4.

At the booking level where after-hours time is neutral, therefore, the session overruns in 65 per cent of cases against 10 per cent at baseline, and the mean overrun has risen from 0.1 to 16.2 minutes. The ninetieth percentile of after-hours work has risen from 21.7 to 69.1 minutes even though the mean is unchanged.

The mechanism is queueing rather than accounting. Overrun depends on the upper tail of the session workload, and adding a patient adds a full draw from a right-skewed distribution to that tail while adding only its mean to the average. Consuming slack therefore degrades reliability faster than it consumes mean time.

The practical consequence is that 'neutral for the clinician' is the wrong test. A clinician whose average after-hours burden is unchanged but who now finishes late in two thirds of sessions, and whose worst days are substantially worse, has not had a neutral change. Mean after-hours minutes is the metric most likely to be reported and the one least able to detect this.

Table 4: Schedule reliability under absorption.

Extra patients	Patients seen	Sessions overrunning	Mean overrun (min)	90th percentile after-hours (min)
0	**11.17**	**1.2%**	**0.1**	**21.7**
1	12.08	5.7%	0.6	23.9
2	13.02	18.8%	2.5	32.0
3	13.94	41.4%	7.5	50.6
4	**14.88**	**65.0%**	**16.2**	**69.1**
5	15.81	82.6%	28.1	87.2
6	16.74	93.0%	42.3	106.2

3.4 Three Ways to Spend the Same Saving

Table 5 and Figure 5 place the scenarios side by side. All three use an identical technical saving and differ only in a scheduling decision.

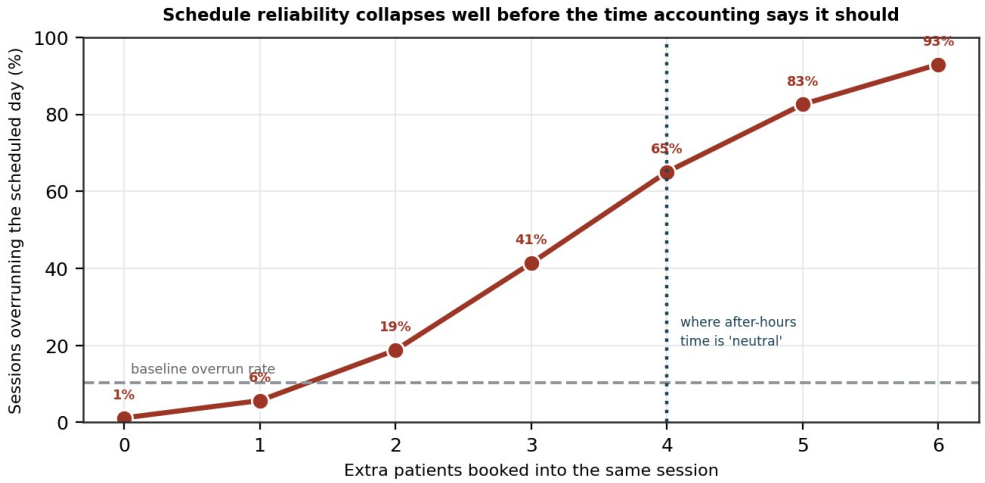


Fig.4: Overrun rate against the number of extra patients booked.

Directing the saving to the clinician gives 11.17 patients, 18.0 after-hours minutes and 1.2 per cent overruns. Directing it to throughput gives 14.88 patients, 40.1 minutes and 65.0 per cent. Splitting it gives 13.02 patients, 23.4 minutes and 18.8 per cent.

The split option is the one we would draw attention to. It delivers most of the throughput gain available at the neutral point while keeping after-hours time well below baseline and the overrun rate close to it. Where the gain is large enough to be worth dividing, an intermediate position dominates either extreme on the combined set of measures.

Table 5: Scenario comparison.

Scenario	Patients seen	After-hours minutes	Total work minutes	Overrun
No tool (baseline)	11.17	37.0	245.2	10.2%
All saving to the clinician	11.17	18.0	205.4	1.2%
All saving to throughput	14.88	40.1	273.8	65.0%
Saving split	**13.02**	**23.4**	**239.4**	**18.8%**

3.5 The Safe Dividend Is a Fraction of the Arithmetic One

Table 6 and Figure 6 ask how many extra patients can be booked while holding both after-hours time and the overrun rate at or below their baseline values.

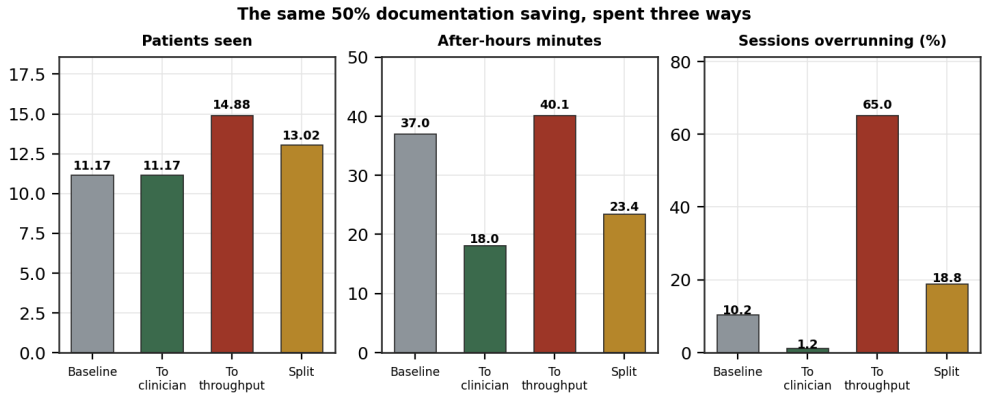


Fig.5: The same 50 per cent documentation saving, spent three ways.

At a 50 per cent documentation reduction the answer is 1 extra patient, taking the clinic from 11.17 to 12.10 patients per session, an increase of 8.3 per cent. Dividing the freed minutes by the mean total work per patient suggests approximately 1. Even a 70 per cent reduction permitted only 2 extra patients on this criterion.

The gap between the two figures is the cost of variability, and it is large: the safe dividend is roughly a quarter of the arithmetic one. A capacity plan built on mean times will book to a level at which the schedule fails routinely, and the failure will present as clinicians running late rather than as a planning error.

We would note that this criterion is strict. An organisation may reasonably accept some deterioration in overrun rate in exchange for access, particularly where waiting times are long. The point is that the trade should be made deliberately and with the overrun consequence quantified, rather than arrived at by dividing minutes.

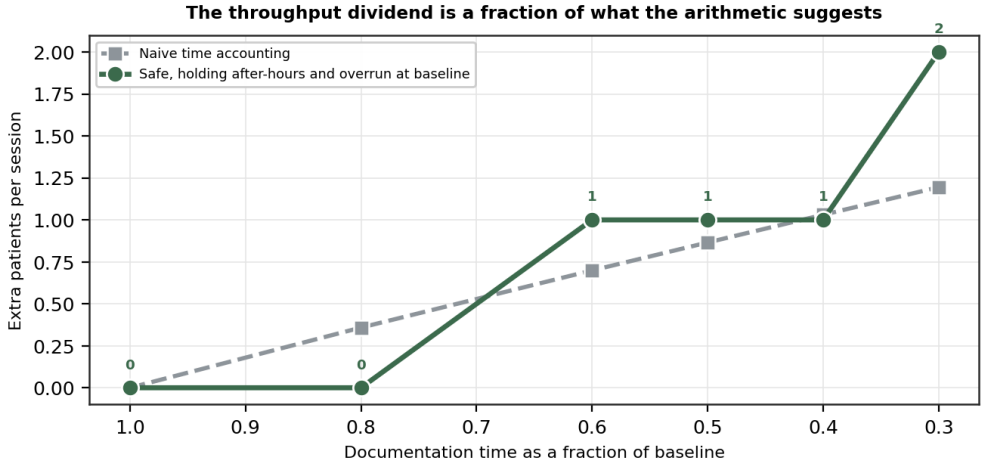


Fig.6: Extra patients permitted by mean-time arithmetic and by holding both after-hours time and overrun at baseline.

Table 6: Safe absorption by documentation reduction.

Tool factor	Extra patients permitted	Patients seen	After-hours minutes	Overrun
1.00	0	11.17	37.0	10.2%
0.80	0	11.17	29.1	4.7%
0.60	1	12.10	24.3	9.2%
0.50	**1**	**12.10**	**20.1**	**6.2%**
0.40	1	12.10	15.9	3.9%
0.30	2	13.02	13.6	9.5%

Criterion: after-hours minutes and overrun rate both at or below their baseline values.

4. DISCUSSION

4.1 Principal Findings

Four findings follow. With the schedule unchanged, halving documentation time reduced after-hours work by 51 per cent and overruns from 10.2 to 1.2 per cent. Booking 4 extra patients returned after-hours time to baseline while raising throughput 33 per cent. At that point

the overrun rate had risen from 10 to 65 per cent, so the position was not in fact neutral. And holding both measures at baseline permitted only 1 extra patient against the 1 implied by mean-time arithmetic.

The unifying observation is that the technology determines the size of a saving and nothing else. Its effect on clinician workload, on access and on punctuality is determined by a scheduling decision, and in most deployments that decision is made by default rather than by deliberation.

4.2 Relation to Existing Work

Time-motion and log-based studies have established the scale of documentation burden and its distribution across and beyond the working day [1-3,9]. Early evaluations of ambient documentation tools report reductions in documentation time and in some cases in self-reported burnout [4-6,10], with heterogeneity that has been attributed in part to differences in how clinics responded organisationally [7,11]. The queueing behaviour of appointment systems and the sensitivity of overrun to service-time variability are established in health operations research [12-19].

We are not aware of these two literatures being brought together quantitatively for this specific question. The contribution here is to show that the operational result of a documentation tool is determined by the scheduling response, and that the response most likely to be chosen on time-accounting grounds degrades punctuality well before it exhausts the mean saving.

4.3 Implications

Figure 7 and Table 7 set out the guidance. Two points deserve comment.

Table 7: What each stakeholder should be shown.

Stakeholder	Question they are asking	Metric that answers it	Metric usually reported
Clinician	Will my day be shorter?	After-hours minutes and overrun rate	Documentation minutes saved
Service manager	Can we see more patients?	Safe extra bookings at constant overrun	Freed minutes divided by slot length
Patients	Will I be seen on time?	Overrun rate and clinic finish time	Not reported
Procurement	Did the tool work?	All of the above, before and after	Documentation minutes saved

State the destination of the saving before deployment. A tool procured to reduce clinician burden and deployed into a scheduling system that automatically fills capacity will not reduce clinician burden, and the subsequent evaluation will record that correctly. Deciding and recording the intended destination at the outset costs nothing and makes the evaluation interpretable.

Report overrun rate alongside after-hours minutes. Mean after-hours time is the natural summary and it is precisely the measure that cannot distinguish a genuinely neutral change from one that has traded predictability for volume. Both quantities are recoverable from scheduling and access logs that services already hold [24,25].

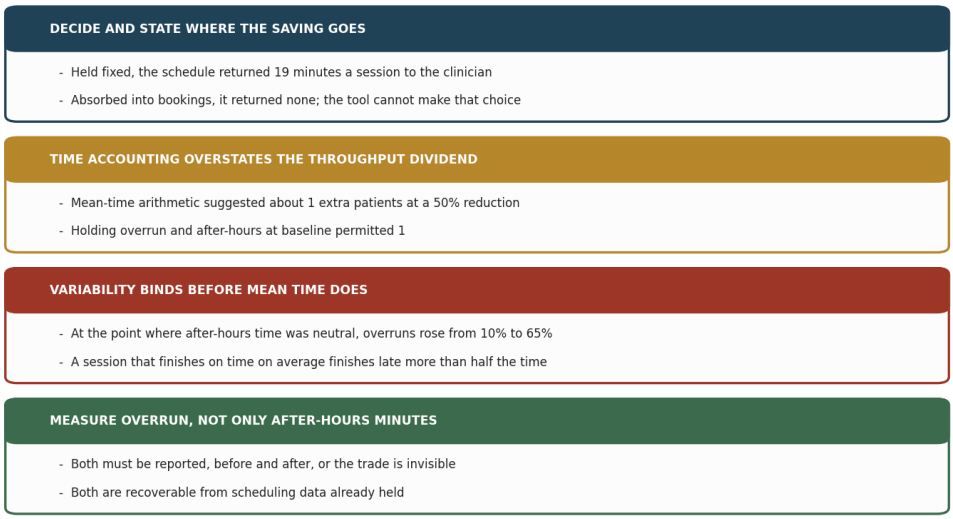


Fig.7: Proposed practice for deploying and evaluating documentation tools.

5. LIMITATIONS

The model is a single clinician in a single session type with independent patients and no interruptions, no urgent add-ons, no shared resources and no carry-over between sessions. Real clinics have all of these, and interruptions in particular are a documented driver of both documentation time and overrun that we have not represented.

Documentation time is reduced by a constant multiplicative factor independent of consultation complexity. Real tools are likely to help more with routine encounters than with complex ones, which would concentrate the saving in the shorter appointments and change the tail behaviour that drives our overrun results.

We model the tool as costless in time. In practice ambient systems require review and correction of generated text, and the quality of that review is a determinant of both time saved and clinical risk. A tool that reduces writing time but adds verification time may have a much smaller net effect than the factor we apply.

Non-attendance is independent and constant, whereas in practice it correlates with appointment type and time of day, and overbooking policies exist precisely to exploit it [19]. An overbooking strategy interacts with everything analysed here and is not represented.

Finally, we treat after-hours minutes and overrun as the outcomes of interest. Neither is burnout, which depends on autonomy, workload composition and much else [20-22], and a reduction in after-hours minutes should not be assumed to produce a proportional improvement in wellbeing.

6. CONCLUSION

Halving documentation time in a simulated clinic reduced after-hours work from 37.0 to 18.0 minutes per session when the schedule was left alone, and reduced it not at all when 4 extra patients were booked to absorb the saving. The technology was identical in both cases.

At the booking level that leaves after-hours time unchanged, the session overran in 65 per cent of cases against 10 per cent at baseline, and the ninetieth percentile of after-hours work rose by 47 minutes despite an unchanged mean. Holding punctuality as well as mean workload at baseline permitted 1 extra patient rather than 1.

The destination of the saving is a decision, not a property of the tool, and the safe size of the throughput dividend is set by variability rather than by mean time. Both are knowable before deployment from data services already hold, and both are usually settled by default instead.

DECLARATIONS

Ethics approval and consent to participate. Not applicable. This study used entirely synthetic data generated by the author. No human participants, clinicians, patients or records were involved, and no service data were accessed. Institutional review board or ethics committee approval and informed consent were therefore not required.

Consent for publication. Not applicable.

Data availability statement. This study used only synthetic data generated by simulation; no real-world, clinical or patient data were used. All results were produced by a simulation script developed by the author in Python using NumPy, SciPy and Matplotlib [28-30], with fixed random seeds. Every model parameter is reported in full in Table 1, and all figures and all values in Tables 2 to 6 can be reproduced exactly from these parameters. The simulation script is available from the corresponding author upon reasonable request.

Conflict of interest. The author declares no competing interests, financial or non-financial, in relation to the subject matter of this manuscript. The author has no financial or advisory relationship with any vendor of ambient documentation, speech recognition or scheduling software.

Funding. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Author's contribution. Dr. Shahida T is the sole author of this work and was responsible for the conception and design of the study, development and implementation of the simulation model, analysis and verification of all reported values, preparation of the figures and tables, formulation of the recommendations, and drafting, revision and final approval of the manuscript. The author is accountable for all aspects of the work.

Acknowledgements. None declared.

REFERENCES

1. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Ann Intern Med. 2016;165(11):753-60.
2. Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan WJ, Sinsky CA, et al. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. Ann Fam Med. 2017;15(5):419-26.
3. Melnick ER, Dyrbye LN, Sinsky CA, Trockel M, West CP, Nedelec L, et al. The association between perceived electronic health record usability and professional burnout among US physicians. Mayo Clin Proc. 2020;95(3):476-87.
4. Tierney AA, Gayre G, Hoberman B, Mattern B, Ballesca M, Kipnis P, et al. Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. NEJM Catal Innov Care Deliv. 2024;5(3).
5. Haberle T, Cleveland C, Snow GL, Barber C, Stookey N, Thornock C, et al. The impact of nuance DAX ambient listening AI documentation: a cohort study. J Am Med Inform Assoc. 2024;31(4):975-79.

6. Shah SJ, Devon-Sand A, Ma SP, Jeong Y, Crowell T, Smith M, et al. Ambient artificial intelligence scribes: physician burnout and perspectives on usability and documentation burden. J Am Med Inform Assoc. 2025;32(2):375-80.
7. Sinsky CA, Rule A, Cohen G, Arndt BG, Shanafelt TD, Sharp CD, et al. Metrics for assessing physician activity using electronic health record log data. J Am Med Inform Assoc. 2020;27(4):639-43.
8. Rotenstein L, Melnick ER, Sinsky CA. Making sense of the EHR burden. JAMA Intern Med. 2024;184(7):715-16.
9. Overhage JM, McCallie D. Physician time spent using the electronic health record during outpatient encounters. Ann Intern Med. 2020;172(3):169-74.
10. Garcia P, Ma SP, Shah S, Smith M, Jeong Y, Devon-Sand A, et al. Artificial intelligence-generated draft replies to patient inbox messages. JAMA Netw Open. 2024;7(3):e243201.
11. Melnick ER, Ong SY, Fong A, Socrates V, Ratwani RM, Nath B, et al. Characterizing physician EHR use with vendor derived data: a feasibility study and cross-sectional analysis. J Am Med Inform Assoc. 2021;28(7):1383-92.
12. Cayirli T, Veral E. Outpatient scheduling in health care: a review of literature. Prod Oper Manag. 2003;12(4):519-49.
13. Gupta D, Denton B. Appointment scheduling in health care: challenges and opportunities. IIE Trans. 2008;40(9):800-19.
14. Ahmadi-Javid A, Jalali Z, Klassen KJ. Outpatient appointment systems in healthcare: a review of optimization studies. Eur J Oper Res. 2017;258(1):3-34.
15. Klassen KJ, Yoogalingam R. Improving performance in outpatient appointment services with a simulation optimization approach. Prod Oper Manag. 2009;18(4):447-58.
16. Robinson LW, Chen RR. Scheduling doctors' appointments: optimal and empirically-based heuristic policies. IIE Trans. 2003;35(3):295-307.
17. Denton B, Gupta D. A sequential bounding approach for optimal appointment scheduling. IIE Trans. 2003;35(11):1003-16.
18. Cayirli T, Veral E, Rosen H. Designing appointment scheduling systems for ambulatory care services. Health Care Manag Sci. 2006;9(1):47-58.
19. LaGanga LR, Lawrence SR. Clinic overbooking to improve patient access and increase provider productivity. Decis Sci. 2007;38(2):251-76.
20. Shanafelt TD, West CP, Dyrbye LN, Trockel M, Tutty M, Wang H, et al. Changes in burnout and satisfaction with work-life integration in physicians during the first 2 years of the COVID-19 pandemic. Mayo Clin Proc. 2022;97(12):2248-58.
21. West CP, Dyrbye LN, Shanafelt TD. Physician burnout: contributors, consequences and solutions. J Intern Med. 2018;283(6):516-29.
22. Gardner RL, Cooper E, Haskell J, Harris DA, Poplau S, Kroth PJ, et al. Physician stress and burnout: the impact of health information technology. J Am Med Inform Assoc. 2019;26(2):106-14.
23. Adler-Milstein J, Zhao W, Willard-Grace R, Knox M, Grumbach K. Electronic health records and burnout: time spent on the electronic health record after hours and message volume associated with exhaustion but not with cynicism among primary care clinicians. J Am Med Inform Assoc. 2020;27(4):531-38.
24. Rule A, Melnick ER, Apathy NC. Using event logs to observe interactions with electronic health records: an updated scoping review shows increasing use of vendor-derived measures. J Am Med Inform Assoc. 2023;30(1):144-54.
25. Baxter SL, Apathy NC, Cross DA, Sinsky C, Hribar MR. Measures of electronic health record use in outpatient settings across vendors. J Am Med Inform Assoc. 2021;28(5):955-59.

26. Micek MA, Arndt B, Tuan WJ, Trowbridge B, Dean SM, Lochner J, et al. Physician burnout and timing of electronic health record use. ACI Open. 2020;4(1):e1-e8.
27. Holmgren AJ, Downing NL, Bates DW, Shanafelt TD, Milstein A, Sharp CD, et al. Assessment of electronic health record use between US and non-US health systems. JAMA Intern Med. 2021;181(2):251-59.
28. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585(7825):357-62.
29. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261-72.
30. Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng. 2007;9(3):90-95.