ORIGINAL RESEARCH • RECHERCHE ORIGINALE

ADMINISTRATION

Prospective time study derivation of emergency physician workload predictors

Grant D. Innes, MD; Robert Stenstrom, MD, PhD; Eric Grafstein, MD; James M. Christenson, MD

ABSTRACT

Background: A reliable emergency department (ED) workload measurement tool would provide a method of quantifying clinical productivity for performance evaluation and physician incentive programs; it would enable health administrators to measure ED outputs; and it could provide the basis for an equitable formula to estimate ED physician staffing requirements. Our objectives were to identify predictors that correlate with physician time needed to treat patients and to develop a multivariable model to predict physician workload.

Methods: During 31 day, evening, night and weekend shifts, a research assistant (RA) shadowed 20 emergency physicians, documenting time spent performing clinical and non-clinical functions for 585 patient visits. The RA recorded key predictors including patient gender, age, vital signs and Glasgow Coma Scale (GCS) score, and the mode of arrival, triage level assigned, comorbidity and procedures performed. Multiple linear regression was used to describe the associations between predictor variables and total physician time per patient visit (TPPV), and to derive an equation for physician workload. Model derivation was based on 16 shifts and 314 patient visits; model validation was based on 15 shifts and 271 additional patient visits.

Results: The strongest predictor variables were: procedure required, triage level, arrival by ambulance, GCS, age, any comorbidity, and number of prior visits. The derived regression equation is: TPPV = 29.7 + 8.6 (procedure required [Yes]) − 3.8 (triage level [1–5]) + 7.1 (ambulance arrival) − 1.1 (GCS [3–15]) + 0.1 (age in years) − 0.05 (n of previous visits) + 3.1 (any comorbidity). This model predicted 31.3% of the variance in physician TPPV (F [12, 29] = 13.2; $p < 0.0001$).

Conclusions: This study clarifies important determinants of emergency physician workload. If validated in other settings, the predictive formula derived and internally validated here is a potential alternative to current simplistic models based solely on patient volume and perceived acuity. An evidence-based workload estimation tool like that described here could facilitate ED productivity measurement, benchmarking, physician performance evaluation, and provide the substrate for an equitable formula to estimate ED physician staffing requirements.

Key words: manpower; workload; emergency; remuneration; staffing

RÉSUMÉ

Contexte : Un outil fiable de mesure de la charge de travail au service d'urgence permettrait de quantifier la productivité clinique aux fins de l'évaluation du rendement et des programmes d'incitations aux médecins, permettrait aux administrateurs de la santé de mesurer la production dans les services d'urgence et servirait de base à l'établissement d'une formule équitable de calcul

des besoins en médecins au service d'urgence. Nous voulions définir des prédicteurs reliés au temps de médecin nécessaire pour traiter des patients et créer un modèle à variables multiples afin de prédire la charge de travail des médecins.

Méthodes : Pendant 31 quarts de jour, de soir, de nuit et de fin de semaine, un attaché de recherche a suivi 20 médecins à l'urgence et a documenté le temps qu'ils ont passé à s'acquitter de fonctions cliniques et non cliniques dans le cas des 585 visites de patients. L'attaché de recherche a consigné des prédicteurs clés, y compris le sexe et l'âge du patient, ses signes vitaux et son résultat sur l'Échelle de Glasgow, ainsi que le mode de transport à l'arrivée, le niveau du triage, la présence d'une comorbidité et les interventions pratiquées. On a utilisé une régression linéaire multiple pour décrire les liens entre les variables des prédicteurs et le temps total de médecin par visite de patient (TPPV) et pour dériver une équation sur la charge de travail des médecins. Le modèle dérivé était fondé sur 16 quarts de travail et 314 visites de patients et sa validation a reposé sur 15 quarts de travail et 271 visites supplémentaires.

Résultats : Les variables des prédicteurs les plus solides étaient l'intervention requise, le niveau du triage, l'arrivée par ambulance, l'échelle de Glasgow, l'âge, la présence d'une comorbidité et le nombre de visites antérieures. L'équation dérivée sur la régression est la suivante : TPPV = 29,7 + 8,6 (intervention requise [oui]) – 3,8 (niveau de triage [1–5]) + 7,1 (arrivée en ambulance) – 1,1 (échelle de Glasgow [3–15]) + 0,1 (âge en années) – 0,05 (n de visites antérieures) + 3,1 (toute comorbidité). Ce modèle a prédit 31,3 % de la variation de la TPPV des médecins (F [12, 29] = 13,2; p < 0.0001).

Conclusions : Cette étude clarifie des déterminants importants de la charge de travail des médecins urgentologues. Si on la valide dans d'autres contextes, la formule de prédiction dérivée et validée ici à l'interne pourrait constituer une solution de rechange au modèle simpliste actuel qui repose uniquement sur le nombre des patients et la gravité perçue de leur cas. Un outil factuel d'estimation de la charge de travail comme celui que l'on décrit ici faciliterait la mesure de la productivité des services d'urgence, les comparaisons, l'évaluation du rendement des médecins et servirait de base nécessaire à l'établissement d'une formule équitable pour évaluer les besoins en médecins dans les services d'urgence.

Introduction

In 2003, the British Columbia Ministry of Health devised a simple emergency department (ED) staffing model, allocating one physician full time equivalent (FTE) per 3000 patients for "high acuity" departments and 1 FTE per 3500 patients in "moderate acuity" departments. In similar fashion, the British Association for Emergency Medicine has proposed a simple formula for estimating ED workforce requirements, defining one workload unit as 3000 patients per physician per year, with adjustments based on whether the cases are "normal, heavy or minor."[1] While patient volume is the primary determinant of physician workload, case mix and complexity are also important, and neither model specifies the factors that define heavier workload. Several authors noted that ambulance patients, referred patients, mental health patients and older patients reflect a more demanding case mix that requires more emergency physician (EP) time per patient.[1–4] Workload is also influenced by sociodemographic factors, site-specific ED processes and available resources (e.g., stretchers, nursing staff), and it is clearly related to procedural requirements, administrative duties, parallel expectations for teaching, documentation and communication with patients, physicians and families.[2,3]

We were unable to find any published literature quantifying the impact of complexity factors on ED workload and time needed to provide physician-related services. In the absence of such information, Britain's National Healthcare System quantifies ED physician workload by dividing the number of emergency visits per annum by the number of doctors.[5] The British Association for Emergency Medicine document[1] suggests an ED with "average" case mix has an admission rate of 15%–20%, with 25% pediatric cases and 50% adult "minor" cases.[1] It is tempting to use triage levels as a measure of workload, but triage levels reflect acuity — not complexity — and interdepartment triage reliability is uncertain and has not been studied.[5–7] In addition, if triage levels are allowed to determine remuneration, gaming may cause triage creep, which will generate unrealistic staffing levels and invalidate important triage and case mix information.

A valid ED workload measurement tool would facilitate ED and physician productivity assessment, and could form the basis for an equitable method of estimating physician staffing needs. Our proxy for workload was EP time required per patient seen. Our primary objective was to identify clinical, demographic and setting-related factors that

correlate with EP workload, and to develop a multivariable linear regression model to predict the amount of physician time required per patient seen in an ED. Our secondary objectives were to validate this model on a separate group of patients, and to identify the relative proportions of EP time consumed by direct patient care, documentation, communication, departmental problem solving, teaching and academic duties.

Methods

Design and setting

This 2-phase prospective cohort study was conducted in the ED of St. Paul's Hospital, an inner city urban teaching centre in Vancouver, Canada. Phase 1 entailed model development ($n = 314$) and Phase 2, model validation ($n = 271$).

Study procedures

A research assistant (RA) was oriented to the clinical and non-clinical tasks of an EP (Fig. 1[8]) and instructed how to gather and record the study variables. The RA was aware the study was a time analysis of EP activity but was blind to the primary objective. During each study shift, the RA shadowed the attending physician for the duration of the shift and measured, with a stopwatch, the time (to the nearest 15 seconds) that the physician spent performing all clinical and non-clinical duties. Using structured data collection forms specific to each individual patient treated, the RA recorded predictor and outcome variables, and EPs documented comorbidity variables for every patient treated. Patients who left before being seen and those not seen by the EP (e.g., direct referrals to other consultants) were not eligible for inclusion.

Performing history and physical
Time spent interviewing the patient and conducting a physical examination. This time includes any reassessments during the ED stay.

Charting and documentation
Time spent by the attending EP documenting findings, results or patient progress on the ED chart.

Ordering tests & treatments
Time spent by the attending EP entering orders for diagnostic tests and treatment modalities.*

Communicating with family and other health professionals
Time spent speaking directly or on the telephone to family members, nurses or other care providers about the patient in question.

Reviewing charts and test results
Time spent reviewing old charts, and reviewing or analyzing results of diagnostic tests performed during the index visit.

Performing procedures
Time spent performing emergency procedures (e.g., airway management, fracture reduction, wound repair, IV access, tube placement).

Providing direct bedside care
Time spent at the patient's bedside directing or supervising patient care (e.g., resuscitation) or administering medications.

Looking up information / references
Time spent looking up medical information in textbooks, journals or online reference sources.

Teaching students or residents[†]
Time spent discussing the patient, the findings or investigations with medical trainees who are involved in the patient's care.

Discussion with patient, and discharge instructions[†]
Time spent explaining treatments, test results, discharge and follow-up instructions with the patient or caregiver.

Other duties
(administrative functions, problem-solving, receiving phone calls)
Time spent performing activities not directly related to the care of the specific patient. Examples include dealing with departmental problems, handling paramedic communications or telephone calls about patients being referred in, assisting in the care of other physicians' patients, and discussions with nurses, managers or other physicians about patient-flow problems.

Chart completion
Time spent after patient discharge completing the chart and recording patient diagnosis, complexity level and procedures performed.

Fig. 1. Definitions of clinical and non-clinical tasks of an emergency physician.
Note: All times refer to time spent by the attending EP. Time spent by medical trainees, nurses or other health providers was not recorded or included in the analysis. *During the study, diagnostic tests were ordered using computerized physician order entry while drug and treatment orders were handwritten on the chart. †In some cases, EPs would discuss findings, treatments and instructions at the bedside with the medical trainee and the patient. In such cases, if it was unclear whether the EP was teaching or providing discharge information, priority was given to the clinical task and time was entered in the category of "Discussion with patient, and discharge instructions." EP = emergency physician; ED = emergency department; IV = intravenous.

Shift selection

EP staffing at the study site involves 49 shifts per week. Of these, 15 (31%) are week-day shifts, 15 (31%) are week-evening shifts, 12 (24%) are weekend day or evening shifts, and 7 (14%) are night shifts. To assure representative patient sampling, a series of study shifts was systematically selected to match this shift distribution as closely as possible and to allow the RA to observe every active physician in the emergency group for at least one shift.

Predictor and outcome variables

Prior to the study, the investigators determined by consensus a set of candidate predictor variables and comorbid conditions likely to influence physician workload and per-patient time requirements (Fig. 2). The primary outcome variable was total physician-time involved in caring for each patient. Secondary outcome variables included the amount of physician time (per patient) spent on the following activities: history and physical exam, charting and documentation, test ordering, communications (with nurses, referring and consulting physicians, other health professionals and families), reviewing charts and test results, performing procedures, direct bedside care, looking up clinical references, teaching students or residents, and other duties (non-physician functions, problem-solving and phone calls). Key patient and utilization outcomes, including time to physician assessment, consultation and lengths of stay were also tracked.

Patient demographics
Age, gender, housing status.

Clinical factors
Mode of arrival, time of arrival, CTAS triage level (I–V), language or communication barrier, presenting vital signs, GCS score, number of medications, disposition (admitted v. discharged), ED procedures performed.

Pre-defined comorbid conditions
Chronic renal disease causing daily symptoms, chronic respiratory disease affecting daily life, chronic cardiovascular illness causing daily symptoms, neurological disability causing daily symptoms, diabetes mellitus, HIV or AIDS, alcohol or drug dependency affecting daily life.

Setting variables
Daily census, physician coverage hours, ED stretcher availability (i.e., overcrowding), seen primarily by house staff, physician years of experience.

Fig. 2. Candidate predictor variables.
CTAS = *Canadian Emergency Department Triage and Acuity Scale*;[11] GCS = Glasgow Coma Scale; ED = emergency department.

Sample size

The unit of observation was the patient visit, and the dependent variable was total physician time per patient visit (TPPV). The sample size calculation was based on a multiple linear regression model (primary objective) using the following assumptions: a minimum of 15 observations per independent variable for the purpose of model stability; a total of 12 candidate independent variables; a β-error of 0.2 (power = 80%) and α-error of 5%; a clinically important effect size (minimum proportion of variance explained by the model [R-squared] divided the proportion of unexplained [error] variance in TPPV to be detected by the regression model of 0.2/0.8 = 0.25. This is considered a moderate to large effect size.[9] Based on these assumptions, 193 patient visits per phase of the study were required (Power and Precision™ Software [Biostat Inc.]).

Statistical analysis

Data were collated and entered into an Excel spreadsheet. Statistical analyses were performed using Statistica and SAS statistical software. Descriptive statistics, including means, medians and standard deviations were determined where appropriate.

Primary objective: To determine the relative impact of the pre-defined candidate predictor variables on physician workload, a multivariable linear regression model was developed, with the dependent variable being total physician time spent with each patient. The usual assumptions underlying linear regression (homoscedasticity of the variance of Y [physician time] at each X [predictor] value, and normality of the distribution of Y at different values of X) were assessed.[9]

The multivariable linear regression model was developed using a forward stepwise selection procedure (F-to-enter = 0.05). Candidate variables that exhibited co-linearity (a high degree of correlation) with each other underwent regression diagnostics. Nonlinear relationships between [age and number of previous visits] with total physician time were assessed using 2nd- and 3rd-order polynomials of these variables.

Model validation: Because predictive statistical models seldom perform as well on subsequent groups of patients as in the sample from which they were derived, the predictor variables and multiple regression model were validated on a second sample of approximately the same size (sample 2). To assure model validity, we assured that both sets passed all regression diagnostics and we assessed the cross-validation R^2 shrinkage between sample 1 and sam-

ple 2. Cross validation shrinkage refers to the difference in the R^2 value between the derivation and validation sets, and experts suggest that shrinkage values less than 0.2 indicate a reliable model.[10]

Ethical considerations

The study involved no experimental treatments and posed no risk to patients. Research assistants were instructed not to observe the physical exam, so there was no breach of patient privacy. No information beyond standard medical assessment was required, and the data collection process did not change the course or duration of the patient's ED visit. Personal identifiers were deleted from the study database, and patients were assigned a unique numeric identifier. No patient information was released to third parties or published in any format. Based on these facts, the Providence Health Care–University of British Columbia Research Ethics Committee approved the study without the need for patient consent. All EPs provided verbal consent for observation, time measurement and data collection.

Results

Over a 6-week period in July and August 2003, the RA covered 31 ED shifts, including 11 (35%) weekday shifts, 11 (35%) week-evening shifts, 6 (19%) weekend day or evening shifts, and 3 (10%) night shifts. The overall study data reflect the treatment of 585 consecutive patients by 20 different EPs. Assumptions underlying multiple linear regression were valid for these data, and 539 (92%) patients had complete data for all candidate predictor variables. The first 314 patient visits were used for model development and the subsequent 271 encounters for model validation.

Mean physician age was 44.5 years (range 30–62), with 16 males and 4 females. Nine physicians have Canadian College of Family Physicians (CCFP) certification (CCFP-EM), 4 Royal College of Physicians and Surgeons of Canada certification (FRCPC), and 2 American Board of Emergency Medicine certification (ABEM). In addition, 3 had dual ABEM/FRCPC certification and 2 had dual ABEM/CCFP-EM certification. Mean number of years in emergency practice was 14.4 (range 3–30).

Table 1 summarizes baseline characteristics for the study patients. The study sample was predominantly male and relatively young (mean 43.7 yr), with a high prevalence of unstable housing, HIV and alcohol or drug dependency, in keeping with the hospital's inner city location. Half of the study patients fell into *Canadian Emergency Department Triage and Acuity Scale* (CTAS)[11] Emergent and Urgent categories (CTAS I–III), and 28% arrived by ambulance.

These characteristics reflect the department's overall patient mix, suggesting that a representative sample was enrolled.

Table 2 shows the primary outcome variable, total physician TPPV, broken down by task. Mean total physician TPPV treated was 19.2 minutes (standard deviation = 14), with history and physical, and documentation being the most time-consuming activities. The time values in Table 2 reflect means for the overall study sample and not for the relevant subgroup. For example, the mean time shown in Table 2 for performing procedures was 1.7 minutes per patient visit studied ($n = 585$) but 7.65 minutes per patient in the subgroup who actually had a procedure performed ($n = 130$). Similarly, mean teaching time was 1.4 minutes per patient visit studied ($n = 585$) but 5.0 minutes per patient in the subgroup who actually had a trainee involved in their care ($n = 139$).

Table 3 shows the impact of key predictor variables on total physician time required per patient. Time per patient visit was closely related to triage level, increasing from 15.2 minutes in Level V to 40.2 minutes in Level I. Table 3 also shows that EPs spent a mean of 17.4 minutes per patient when trainees were involved in care versus 19.8 minutes per patient when they managed the case alone ($p = 0.09$). Of note, housing status, abnormal vital signs (Y/N),

Table 1. Baseline patient predictor variables (*n* = 581)	
Variable	No. (and %)*
Age, yr; mean (and SD)	43.7 (18.6)
Female gender	219 (37.7)
Unstable housing†	49 (8.4)
Arrival by ambulance	165 (28.4)
CTAS triage category	
Level I–III	279 (48.0)
Level IV–V	302 (52.0)
GCS score <15	85 (14.6)
Systolic BP <90 mm HG (*n* = 489)‡	13 (2.7)
ED procedure performed	130 (22.4)
Chronic comorbid condition	
Renal disease	24 (4.1)
Respiratory disease	31 (5.3)
Cardiovascular illness	37 (6.4)
Neurological disability	18 (3.1)
Diabetes mellitus	26 (4.5)
HIV or AIDS	48 (8.3)
Alcohol/drug dependency	92 (15.8)

*Unless otherwise specified.
†Shelter, hotel or no fixed address.
‡489 patients had vital signs taken.
SD = standard deviation; CTAS = *Canadian Department Emergency Triage and Acuity Scale*; GCS = Glasgow Coma Scale; BP = blood pressure; ED = emergency department;

and time of day seen (0800–2359 v. 0000–0759) were not significantly associated with EP TPPV.

Multivariable analysis showed that the most powerful independent predictors of workload were the need for a procedure, triage level, arrival by ambulance, Glasgow Coma Scale (GCS) score, age, number of previous visits, and presence of any of the pre-defined comorbidities. The regression equation derived from these data are as follows: TPPV = 29.7 + 8.6 (procedure required [Yes]) – 3.8 (triage level I–V) + 7.1 (ambulance arrival) – 1.1 (GCS 3–15) + 0.1 (age in years) – 0.05 (n of previous visits) + 3.1 (any comorbidity). This model predicted 31.3% of the variance in physician TPPV (f [12 291 = 13.2; $p < 0.0001$). When validated on 271 separate patient visits, this model accounted for 28.3% of the variance in total physician TPPV, demonstrating a cross-validation shrinkage R^2 of only 3% (or about 10%, relative to the total mount of variation that the derivation model accounted for). Cross validation shrinkage refers to the difference in the R^2 value between the derivation and validation sets, and experts suggest that shrinkage values less than 20% indicate a reliable model.[3]

Table 4 displays univariable and multivariable correlation (r) values for the model predictors, coming from both the derivation and validation sets. In addition, this table provides estimates of the added time related to each predictor variable, after adjustment for other variables in the model.

Table 2. Mean time performing clinical tasks in 585 consecutive patients

Task	Time in minutes; mean (and SD)	% of total time
History and physical	5.3 (5.7)	27.6
Initial documentation	2.2 (2.0)	11.5
Ordering tests	0.6 (1.0)	3.1
Communicating*	1.6 (1.2)	8.3
Performing procedures†	1.7 (5.1)	8.9
Bedside care	0.7 (3.7)	3.6
Reviewing charts	1.0 (1.9)	5.2
Checking references	0.1 (0.6)	0.5
Teaching†	1.4 (3.4)	7.3
Discharge instructions	1.5 (5.0)	7.8
Other tasks/problems‡	1.8 (3.7)	9.4
Chart completion	1.5 (0.5)	7.8
Total physician time	19.2 (14)	100

*With families and health professionals, about the specific patient.
†Mean time performing procedures was 1.7 minutes per patient visit studied (n = 585) but 7.65 minutes per patient in the subgroup who actually had a procedure performed (n = 130). Mean teaching time was 1.4 minutes per patient visit studied (n = 585) but 5.0 minutes per patient in the subgroup who actually had a trainee involved in their care (n = 139).
‡ Administrative functions, problem-solving, receiving phone calls.

Discussion

This prospective cohort study documented actual physician time spent per patient in a large study cohort during a representative sample of ED shifts. It identified the strongest determinants of EP workload (defined as time spent per patient visit) and incorporated them into a predictive formula that was then validated on a separate set of patients. This model is a potential alternative to simplistic and inequitable workload models that consider only patient volume and perceived acuity. To our knowledge, it is the only existing ED physician workload model derived from actual patient visit and physician activity data.

The strength of these findings is bolstered by the detailed prospective data collection; by the inclusion of representative shifts, consecutive emergency patients and

Table 3. Impact of key predictor variables on total mean physician time per patient visit (n = 581)

Predictor	Minutes (and SD)	No. of patients
Procedure performed		
No	17.0 (12.3)	451
Yes	26.9 (16.7)	130
CTAS level		
I	40.2 (22.7)	8
II	25.3 (20.1)	72
III	21.8 (13.0)	199
IV	15.6 (10.1)	205
V	15.2 (12.6)	97
Arrival by ambulance		
No	16.6 (11.4)	416
Yes	25.8 (17.4)	165
Glasgow Coma Scale score		
15	18.4 (12.6)	496
<15	23.7 (20.0)	85
Age		
<75	18.6 (13.5)	529
≥75	25.5 (17.0)	52
Gender		
Male	18.3 (12.7)	362
Female	20.7 (15.8)	219
Defined comorbid illness		
No	17.4 (12.7)	277
Yes	20.9 (15.0)	304
Trainee involvement		
No	19.8 (13.6)	442
Yes	17.4 (15.3)	139

SD = standard deviation; CTAS = Canadian Emergency Department Triage and Acuity Scale[11]
Note: Minutes are mean total minutes per patient in patients with and without the relevant predictor.

every ED physician working during the study period; by the precise time tracking and analysis of many candidate predictor variables; and by the 2-step validation process used. If validated in diverse ED settings, the key predictors and weightings reported here could form the basis of a common provincial or national ED workload measurement tool. Such a tool would be a more reliable method of quantifying individual physician productivity for performance evaluation and physician incentive programs; it would help health administrators quantify ED outputs at a macro level; and it could provide the basis for an equitable formula to estimate physician staffing requirements in diverse EDs with constantly changing patient complexity.

Productivity measurement and manpower estimation

Emergency physician remuneration is increasingly based on alternate payment plans that specify physician compensation levels and the expected number of working hours per annum. What remains controversial is the number of EPs allocated to staff a given ED. In the face of competing demands from multiple emergency groups, health funders need equitable, transparent allocation models and also need to assure value for their investment. Neither need can be fulfilled without a valid measure of physician productivity.

Patient volume is the default productivity measure, but volume alone does not predict workload or manpower needs, a fact that is increasingly apparent as ED case mixes become more complex. Our premise is that "time needed to provide necessary service" is the key measure of workload — particularly when discussing ED staffing models. In designing the study, we included clinical, administrative, educational and supervisory functions in our model development because all of these are components of workload.[1,12] In 1990, Graff described the need for multivariable complexity assessment and found that a workload formula incorporating volume, patient length of stay, service intensity and service type more accurately estimated the amount of time EPs spent with patients than a volume alone formula.[8,13] Previous authors have noted that mental and physical effort, difficulty, urgency and psychological stress are also important factors.[4,12,14] These are less objective and more difficult to quantify, and we did not incorporate them in our methodology.

A practical workload measurement tool

The key predictors retained in the proposed workload formula, including age, gender, arrival mode, number of previous visits, and CTAS level are already captured in most EDs. Glasgow Coma Scale score can be recorded at triage and incorporated in the ED triage database, and procedures performed are often part of "shadow billing" requirements for contract EDs. This makes automatic electronic workload scoring feasible for each patient visit, and for the ED as a whole. Tracking quantitative workload over time in this manner allows ED directors to more precisely tailor shifting patterns to clinical need. Common data capture mechanisms between hospitals would enable benchmarking and inter-site productivity comparison.

Table 4. Correlations and standardized beta values for key workload predictors

Variable	Derivation sample ($n = 314$)				Validation sample ($n = 271$)			
	Univariable		Multivariable model		Univariable		Multivariable model	
	r	p	β*	95% CI	r	p	β	95% CI
Procedure performed (Y/N)	0.24	<0.001	9.8	6.7 to 12.9	0.36	<0.001	9.4	5.7 to 13.1
CTAS triage level†	−0.29	<0.001	−4.1	−2.5 to −5.7	−0.21	<0.001	−3.2	−1.6 to −4.8
Arrival by ambulance (Y/N)	0.33	<0.001	6.1	2.8 to 9.4	0.27	<0.001	11.3	7.7 to 14.9
GCS score‡	0.14	0.02	2.2	0.7 to 3.7	0.09	0.13	0.1	−1.1 to 1.2
Comorbid condition (Y/N)§	0.17	0.006	3.5	1.7 to 5.3	0.13	0.04	2.8	1.4 to 4.2
Age of patient¶	0.24	<0.001	0.8	0.1 to 1.5	0.19	0.007	0.6	0.3 to 1.4
No. of previous visits by patient**	−0.05	0.41	−0.5	−1.2 to 0.2	−0.005	0.92	−0.7	−0.1 to −1.3

CI = confidence interval; CTAS = *Canadian Emergency Department Triage and Acuity Scale;*[11] GCS = Glasgow Coma Scale
Multivariable values are adjusted for all other variables in the model.
*β refers to standardized beta value. The β value is the point estimate of the time in minutes that each variable adds, relative to the referent category. The 95% CI is for the time in minutes.
†For each one-level increase in CTAS triage level (e.g., from I to II) the workload decreases as quantified in the Table.
‡For each one point the GCS score falls, the workload increases as quantified in the Table.
§See Fig. 2 for a list of pre-defined comorbid conditions.
¶For every 10 years that patient age increases, the workload increases as quantified in the Table. See text under "Statistical considerations" for a discussion of age.
**As the number of the patient's previous visits rises by 10, the workload decreases. Univariable r and associated p are non-significant, but inclusion in the model is based on significance in the multivariable model and significant multivariable beta.

Physician evaluation and incentive systems

Given an established level of physician staffing, increased physician productivity correlates with increased throughput and reduced patient waiting times. To enhance productivity and achieve these objectives, physician incentive programs are increasingly being described in the quality literature and implemented in various medical disciplines. Increased productivity is desirable; however excessively rapid throughput may be associated with medical error, compromised patient–physician interaction, adverse outcomes and patient dissatisfaction; hence the American College of Emergency Physicians established maximum productivity benchmark of 2.5 patients per EP per hour.[13,15–17] Effective incentive systems will increase physician efficiency without jeopardizing other important outcomes, but tracking physician performance and managing incentive programs will depend on the ability to measure workload and productivity in a meaningful way.

ED manpower estimation

The model derived in this study is one possible workload measurement tool that could be used to compare the relative productivity of 2 EDs or 2 EPs but, because of ED patient arrival variability, it cannot be used as an ED staffing formula. To illustrate, many departments have low volume periods (e.g., nights) which, according to the formula, would justify less than a complete EP. Regrettably, EPs come only in integer values, and one is the usual minimum staffing level.

Similarly, in using data from this study to determine staffing needs during higher volume periods (e.g., evenings) it might be tempting to conclude, based on the ~20 minute average TPPV in this sample, that 1 physician could see 24 patients in an 8-hour (480 min) shift, with little or no waiting time. Funding agencies may logically be tempted to multiply the average TPPV by the annual ED census to determine the total number of annual physician hours (and physician FTEs) required. Such a staffing mechanism would enable timely physician assessment for patients — but only if the physician worked continuously, if all arriving patients were of the same (average) complexity, and if there was a constant 19.7-min time interval between each patient registration. In reality, there is wide variability in ED patient arrival rate and complexity. Clearly, if several high acuity patients arrive within minutes of each other, more than one physician is required to provide timely medical response to all of them; hence it is necessary to fund physician "overcapacity" to deal with high volume/high complexity inflow periods. Consequently, if a formula like the one proposed here is used to estimate overall department workload and staffing needs, a correction factor (multiplier) is required to address variability-related concerns.

The required degree of physician "overcapacity" will depend mainly on the level of input variability (which can be described using basic data available in most EDs) and on the tolerance for how many arriving patients can be allowed to wait, and for how long. High input variability and high expectations for physician timeliness (e.g., 90% of Emergent and Urgent patients seen within CTAS time frames) will lead to a relatively large "overcapacity correction factor" and a need for more funded EPs. Less extreme input variability and lower expectations for prompt service would generate a lower "overcapacity correction factor" and fewer funded EPs. These concepts illustrate how less urgent patients can actually enhance ED efficiency and cost-effectiveness by providing a buffer of patients who can safely be queued for longer time periods.

Statistical considerations

This study is based on the use of linear regression for the prediction of total physician TPPV. It may be that the relationship of certain variables with physician time is not linear. For example, age may have a "U-shaped" association with physician time, such that the very young and the very old require more physician time. We investigated this possibility for all of the continuous variables by including higher-order polynomial equations during the model development phase, but these did not improve the predictive ability, so they were excluded from the final models. However, we had few young children and infants in our sample, so these data should not be extrapolated to settings that see a large proportion of children.

The clinical relevance of R^2 in this study is 2-fold. First, it represents the proportion of variance in total physician time that is explained by the predictor variables. In the biologic and social sciences, an R^2 value of 0.3 (30%) is clinically important, and is considered to represent a moderate amount of variance.[9] Cross-validation R^2 shrinkage — difference in the amount of variation in the dependent variable (physician time) that is explained by the same model in a different set of patients — is important because it reflects the stability (or reliability) of the model when tested in a new sample. The shrinkage R^2 between these 2 samples was about 10% of the total variation explained by this model. Shrinkage values less than 20% indicate a reliable model.[4]

Limitations

The primary limitation of this study was that it was per-

formed in one urban hospital with a relatively homogenous group of physicians. Without further study and validation in other settings, the findings cannot be assumed to be generalizable. Time spent examining patients, documenting, performing procedures, communicating and carrying out other clinical tasks will vary from physician to physician and department to department. Correlation (non-independence) of observations that occurs *within* a group of EPs would effectively decrease the number of independent observations and reduce the degrees of freedom for calculated p values, making the true p values for the univariable, multivariable models and R^2 values higher than the estimates presented. However, the unit of analysis for this study is the patient visit (not the physician); hence the estimate of variability in physician time explained by this model (i.e., R^2), and the associated shrinkage R^2 for the 2 samples will not be affected.

A second limitation is that this workload formula considers only ED duties. In smaller community hospitals and non-teaching hospitals, EPs may also provide non-ED services, such as participation on cardiac arrest teams, covering the case room or dealing with patient deterioration on hospital wards. The opinion of the authors is that an ED workload measurement tool should not attempt to incorporate these non-ED services — rather that they should be viewed as distinct from emergency care, and measured and remunerated using other mechanisms. This position is in keeping with recommendations of the American College of Emergency Physicians, the American Academy of Emergency Medicine and the Society of Academic Emergency Medicine, who discourage the use of EPs to provide patient care outside the ED. In addition, rural concerns, such as time associated with patient transfers, were not addressed. Other factors, like availability of support staff, local procedures and ED overcrowding, will all influence the efficiency of staff and the rate that they can see patients. These were not considered.

Another limitation of this study is that it equates workload and time investment. This seems logical but it may be simplistic in that not all work is equally taxing or stressful, and it does not take into account qualitative factors and differing workload intensity related to different tasks, different levels of multi-tasking and different levels of associated stress and risk. Perhaps 30 minutes spent suturing should be considered a "different" workload than 30 minutes interviewing a suicidal patient, although this difference would be difficult to justify in a staffing allocation model.

One of our *a priori* assumptions was that language barriers would increase physician time requirements. Our data suggest otherwise; however, this finding may reflect a flaw in our study methodology. We identified patients' primary language based on the "preferred language" recorded by the registration clerk. Unfortunately, we found little correlation between "preferred language" and "communication barrier." Many patients whose first language was Tagalog or Mandarin spoke fluent English, while many intoxicated, demented or seriously ill "English speakers" did not. Researchers attempting to replicate this work should attempt to identify "communication barrier" rather than "preferred language."

A minor limitation is that we did not document time spent by the "index" physician providing care for other physicians' patients in the ED. Nor did we quantify the care provided by residents and medical students; therefore, although we accurately captured the total time spent by attending EPs providing care and supervising care provided by trainees, we cannot make any conclusions regarding the total care time provided by trainees and attending physicians.

Finally, the Hawthorne effect is a potential factor in this type of study. It is conceivable that an RA with a stopwatch could have motivated a small overall increase in physician work speed and efficiency, hence a slight underestimation of total clinical TPPV. However, because EPs knew only that we were studying factors associated with patient complexity and time demand, and were unaware of the primary objectives and analysis plan, it is unlikely they could have consciously or subconsciously skewed the data in a meaningful way.

Conclusion

This prospective study clarifies important determinants of EP workload. If validated in other settings, the predictive formula derived and internally validated here is a potential alternative to current simplistic models based solely on patient volume and perceived acuity. An evidence-based workload estimation tool like that described here could facilitate ED productivity measurement, benchmarking, physician performance evaluation, and provide the substrate for an equitable formula to estimate ED physician staffing requirements.

Competing interests: None declared.

References
1. British Association for Emergency Medicine. The workforce in emergency medicine. Available: www.baem.org.uk/workforce .doc (accessed 2005 July 23).

2. Weinger MB, Herndon OW, Zornow MH, Paulus MP, Gaba DM, Dallen LT. An objective methodology for task analysis and workload assessment in anesthesia providers. Anesthesiology 1994;80:77-92.

3. Parshuram CS, Dhanani S, Kirsh JA, Cox PN. Fellowship training, workload, fatigue and physical stress: a prospective observational study. CMAJ 2004;170:965-70.

4. Orozco P, Garcia E. The influence of workload on the mental state of the primary health care physician. Fam Pract 1993;10:277-82.

5. NHS Healthcare Commission. Acute hospital portfolio topic guide: accident and emergency — October 2004. Available: www.healthcarecommission.org.uk/assetRoot/04/01/48/51/04014851.pdf (accessed 2005 July 22).

6. Beveridge R, Ducharme J, Janes L, Beaulieu S, Walter S. Reliability of the Canadian Emergency Department Triage and Acuity Scale: inter-rater agreement. Ann Emerg Med 1999;34:155-9.

7. Fernandes CM, Wuerz R, Clark S, Djurdjev O. How reliable is emergency department triage? Ann Emerg Med 1999;34:141-7.

8. Graff LG, Radford MJ. Formula for emergency physician staffing. Am J Emerg Med 1990;8:194-9.

9. Cohen J, Cohen P. Applied multiple regression/correlation analysis for the behavioral sciences. 2nd ed. Hillsdale (NJ): Erbaum and Associates; 1983.

10. Kleinbaum D, Kupper L, Muller K. Applied regression and other multivariable methods. 2nd ed. Boston (MA): PWS-Kent Publishing; 1988.

11. Beveridge R, Clarke B, Janes L, Savage N, Thompson J, Dodd G, et al. Canadian Emergency Department Triage and Acuity Scale: implementation guidelines. Can J Emerg Med 1999;1(3 suppl). Online version available at: www.caep.ca/002.policies/002-02.ctas.htm.

12. Chisholm CD. Collison, EK. Nelson, DR. Cordell WH. Emergency department workplace interruptions: Are emergency physicians "interruptdriven" and "multitasking"? Acad Emerg Med 2000;7:1239-43.

13. Graff LG, Wolf S, Dinwoodie R, Buono D, Mucci D. Emergency physician workload: a time study. Ann Emerg Med 1993;22:1156-63.

14. Bertram DA, Hershey CO, Opila DA, Quirin O. A measure of physician mental work load in internal medicine ambulatory care clinics. Med Care 1990;28:458-67.

15. Endsley S, Kirkegaard M, Baker G, Murcko AC. Getting rewards for your results: pay-for-performance programs. Fam Pract Manag 2004;11(3):45-50.

16. Khan NS, Simon HK. Physician productivity — Can it be enhanced without impacting patient satisfaction and turnaround times? [abstract]. Acad Emerg Med 1999;6(5):401.

17. American Academy of Emergency Medicine. Position statement on emergency physician-to-patient ED staffing ratios. 2001 Feb 22. Available: www.AAEM.org.

Correspondence to: Dr. Grant Innes, Chairman, Department of Emergency Medicine, St. Paul's Hospital, 1081 Burrard St., Vancouver BC V6Z 1Y6; ginnes2@providencehealth.bc.ca