Predicting the Cost and Pace of Pharmacogenomic Advances: An Evidence-Based Study

Ramy Arnaout,[1,2,3*] Thomas P. Buck,[1,3] Paulvalery Roulette,[3†] and Vikas P. Sukhatme[2,3]

BACKGROUND: Adverse outcomes associated with prescription drug use are common and costly. Many adverse outcomes can be avoided through pharmacogenomics: choosing and dosing of existing drugs according to a person's genomic variants. Finding and validating associations between outcomes and genomic variants and developing guidelines for avoiding drug-related adverse outcomes will require further research; however, no data-driven estimates yet exist for the time or money required for completing this research.

METHODS: We identified examples of associations between adverse outcomes and genomic variants. We used these examples to estimate the time and money required to identify and confirm other associations, including the cost of failures, and to develop and validate pharmacogenomic dosing guidelines for them. We built a Monte Carlo model to estimate the time and financial costs required to cut the overall rate of drug-related adverse outcomes by meaningful amounts. We analyzed the model's predictions for a broad range of assumptions.

RESULTS AND CONCLUSIONS: Our model projected that the development of guidelines capable of cutting overall drug-related adverse outcomes by 25%–50% with current approaches will require investment of single-digit billions of dollars and take 20 years. The model forecasts a pump-priming phase of 5–7 years, which would require expenditures of hundreds of millions of dollars, with little apparent return on investment. The single most important parameter was the extent to which genomic variants cause adverse outcomes. The size of the labor force was not a limiting factor. A "50 000 Pharmacogenomes Project" could speed progress. Our approach provides a template for other areas of genomic research.

© 2012 American Association for Clinical Chemistry

Genomic research is widely expected to transform medicine but progress has been slower than expected (1, 2). To critics, delays represent broken promises and a sign that at least some of the money spent on genomic research might be better spent elsewhere (3–5). To proponents, the pace simply underscores the complexity of the relationship between genetics and disease and argues, if anything, for increased funding (6, 7). Thus far, these competing points of view have been based on qualitative predictions. An alternative basis is quantitative modeling. Quantitative modeling is already used to inform opinions on everything from presidential elections to the weather (8, 9). Through the use of technological forecasting methods (10, 11), modeling can similarly be applied to genomics to help set expectations and inform investment and other decisions.

Within genomics an area that has begun to affect clinical practice is pharmacogenomics: choosing and dosing of prescription drugs according to a person's genomic variants (12, 13). Pharmacogenomics has a special bearing on clinical chemistry and molecular pathology, because these services measure drug levels and perform diagnostic genetic tests (14). Years of research and growing clinical experience provide data for modeling in pharmacogenomics, and problems related to prescription drugs are common. For these reasons, we chose pharmacogenomics as our area of investigation.

Prescription drugs are the mainstay of modern medicine but often fail to work as intended. For example, of the 30 million Americans who take aspirin to prevent stroke or heart disease, one quarter experience treatment failure in the form of aspirin resistance (15, 16); of 60 million who take statins to lower cholesterol, 3 million experience muscle pain, increased liver enzymes, or rhabdomyolysis as side effects that can lead to nonadherence and thereby to problems related to hypercholesterolemia (16, 17). Treatment failure and side effects lead to adverse drug events, which are defined as injuries due to drug-related medical inter-

[1] Departments of Pathology and [2] Medicine, Beth Israel Deaconess Medical Center, Boston, MA; [3] Harvard Medical School, Boston, MA.
* Address correspondence to this author at: Beth Israel Deaconess Medical Center, 330 Brookline Ave., BIDMC East/CLS 741, Boston, MA 02215. Fax

vention *(18)*. These events affect 1 in 5 outpatients and 1 in 11 inpatients *(19)* and they $80 billion per year in the US alone *(20)*. Lowering rates of treatment failure and side effects—collectively, drug-related adverse outcomes—is thus a meaningful target for clinical improvement *(6, 21–25)*.

Preventable causes of adverse events—drug nonadherence, drug–drug interactions, and medical error—constitute only a fraction of the total *(19)*. Of the 80% of outpatient adverse drug events and nearly half of the inpatient events currently considered nonpreventable, a majority are thought to be caused by genomic variation *(19, 26, 27)*. For example, many patients on warfarin, an anticoagulant used by 30 million Americans, experience unintended bleeding because of the variation among individuals in dosing requirements *(28, 29)*. Two thirds of this variability can be explained by single-nucleotide polymorphisms in a set of 6 genes; the total percent attributable risk estimated for this adverse outcome is 77% *(26)*. Genetic screening for variants in genes of the major histocompatibility complex in HIV-positive patients taking the reverse-transcriptase inhibitor abacavir almost completely avoids potentially life-threatening hypersensitivity reactions *(30)*.

Conceptually, modeling the investment required to learn how to avoid adverse outcomes is straightforward, as in the following equation:

$$Total\ cost = (Cost\ per\ association)$$
$$\times (Number\ of\ associations\ per\ adverse\ outcome)$$
$$\times (Number\ of\ adverse\ outcomes\ per\ drug)$$
$$\times (Number\ of\ drugs).$$

In practice, care must be taken not to oversimplify by ignoring: (*a*) variability in the cost per association, the number of associations per adverse outcome, and so on; (*b*) limitations imposed by the percent attributable risk due genomics; and (*c*) potentially hidden factors such as the cost of research failures. In modeling the amount of time required, similar care must be taken to account for potential limitations imposed by, for example, the size of the available work force. Monte Carlo models address such variability by sampling at random from the raw data—e.g., from the set of costs per association for different associations—instead of using averages. Other pitfalls can be avoided through careful compilation of data and choosing realistic goals to model—e.g., reducing the rate of adverse outcomes by less than the maximum percentage attributable to genomics.

In pharmacogenomics, several associations between germline genetic variants and adverse outcomes are in advanced stages of clinical investigation. For these associations, the incidence of the adverse outcome, the frequency of the associated variant in clinical populations, and the percent attributable risk of each variant for its associated adverse outcome are all known. To build a simple Monte Carlo model for discovering additional associations, we used these data, the timeline of the associations' discovery and confirmation, the cost of this work, and the number of people taking the most common prescription drugs. We used this approach to model the time and financial resources required to cut the overall incidence of drug-related adverse outcomes by meaningful amounts—e.g., by half.

Methods

MODEL

Conceptually, the model follows the outline of the equation presented above. We considered the broad class of associations that involve germline, as opposed to somatic (tumor), bacterial, or viral, genetic variants. Following standard requirements for translation of biomedical discoveries into clinical use *(6, 31)*, we took the starting and end points of the process to be the discovery of relevant variants and the validation of the utility of clinical guidelines, respectively. The process consisted of 4 steps: discovery of a candidate association, confirmation of the association in a large clinical cohort, trialing a guideline for using the association to decrease the incidence of the adverse outcome, and confirming the guideline's utility in a large clinical cohort. Upstream discoveries (e.g., discovery of the cytochrome P450 system) and infrastructural investments (such as contributions to the SNP Consortium) were considered outside the process, because the money for them has already been allocated or spent.

As data for our model, we selected 8 associations involving 6 prescription drugs (clopidogrel, warfarin, escitalopram, carbamazepine, the nicotine-replacement patch, and abacavir) and 1 drug class, the statin class of anticholesterol drugs (hydroxymethylglutaryl-CoA reductase inhibitors) represented by atorvastatin, simvastatin, and pravastatin (Table 1; see Table S1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol59/issue4). We initially based our selection of associations on strength of evidence *(32)* and the number of patients who use the drug (see Supplementary Methods in the online Data Supplement). We then modified the list to ensure representation of a range of drug indications and mechanisms. For each association, we searched Medline for primary reports that documented each step of the process described above (see Table S1 and Supplemental References in the online Data Supplement). When multiple reports described substantially similar results in the same year, we con-

Drug	Indication	Adverse outcome (frequency)	Association (frequency)	Percent attributable risk	References
			Table 1. Genetic associations for drug-related adverse outcomes.[a]		
Clopidogrel	Acute coronary syndrome	Cardiovascular event (0.08)	CYP2C19*2 (0.28)	29%	Hulot et al. (41)
Statins[b]	High cholesterol	Myopathy (0.05)	SLCO1B1 (rs4363657 and rs4149056, 0.28)	60%	Link et al. (42)
Escitalopram	Depression	Treatment failure (0.37)	HTR2A (rs2224721, 0.50; rs9316233, 0.55)	2.3%	Uher et al. (43), http://www.snpedia.com
Nicotine patch	Smoking	Return to smoking within 12–16 weeks (0.85)	COMT Met/Met homozygote (0.31)	9.7%[c]	Johnstone et al. (44), Munafo et al. (45)
Warfarin	Clotting risk	Bleeding (0.10), dose out of range (0.40[d])	VKORC1 (0.64)	30%[e]	Wadelius et al. (46), Kangelaris et al. (47), Aithal et al. (48)
			CYP2C9*2 and *3 (0.34)	12%[e]	Wadelius et al. (46), Kangelaris et al. (47), Aithal et al. (48)
Abacavir	HIV infection	Hypersensitivity reaction (0.03)	HLA-B*5701 (0.056)	100%	Mallal et al. (30)
Carbamazepine	Seizures	Hypersensitivity reaction (0.04)	HLA-B*1502 (0.077)	100%	Chen et al. (49)

[a] Our model was built on the basis of well-studied associations between genetic variants and drug-related adverse outcomes. Associations were selected to represent a range of indications, drug targets, and adverse outcomes that best represent the breadth of prescription drugs in use today. "rs" denotes a single-nucleotide polymorphism in standard nomenclature.
[b] Hydroxymethylglutaryl-CoA reductase inhibitors atorvastatin, pravastatin, and simvastatin.
[c] Mean; range, 8.6%–10.7%.
[d] Mean; range, 0.30–0.50.
[e] For dose out of range.

sidered both reports for each step. In cases for which no clinical validation existed, we included no report for that step. From government-compiled salary and other financial information and bibliometric analysis (see Supplementary Methods and Fig. S1 in the online Data Supplement), we calculated the total cost of producing each report (including salary, fringe benefits, indirect costs, and supplies) as being proportional to the number of coauthors (mean of $35 000–$50 000 per coauthor's contribution). When hundreds of collaborators were named besides the coauthors, we counted only the authors. Note that because we calculated the cost of a coauthor's contribution by first calculating the cost of a coauthor per year and then dividing by that coauthor's productivity in terms of published reports, the cost of "failure"—research that does not lead to published work or leads to work that does not confirm findings—is included in this cost (see Supplementary Methods in the online Data Supplement).

We estimated the time required for each step as the number of years between reports for each step (see Table S2 in the online Data Supplement). We estimated the time required for the first step as the median number of years since the most recent pertinent research reports cited by the first-step report (defined as those cited in the introduction of the first report; see Table S3

in the online Data Supplement). In the 2 cases in which named clinical trials completed enrollment in 2011, we set a date of 2012 for the last step of the process we have described above (see legend to Table S2 in the online Data Supplement). As with the cost of failure, this method includes the time spent performing "nonproductive" research.

Because initial and smaller reports tend to overstate effects (33), we used the confirmatory report (or a subsequent metaanalysis, when available) to record each association's percent attributable risk for its adverse outcome (Table 1). For each drug, we used product inserts, US Food and Drug Administration safety information, and Medline to determine its adverse-outcome profile, which was defined as the frequencies of treatment failure and side effects documented to lead to drug nonadherence or discontinuation (Fig. 1; see Table S4 and Fig. S3 in the online Data Supplement). We considered 60% as a conservative estimate of the maximum incidence that genetic guidelines can prevent, because this percentage is substantially less than the 83% of outpatient adverse drug events considered nonpreventable (19) and possibly genetic and is less than the 77% of the warfarin dose variation thought to be due to genetic causes (26). Nevertheless,
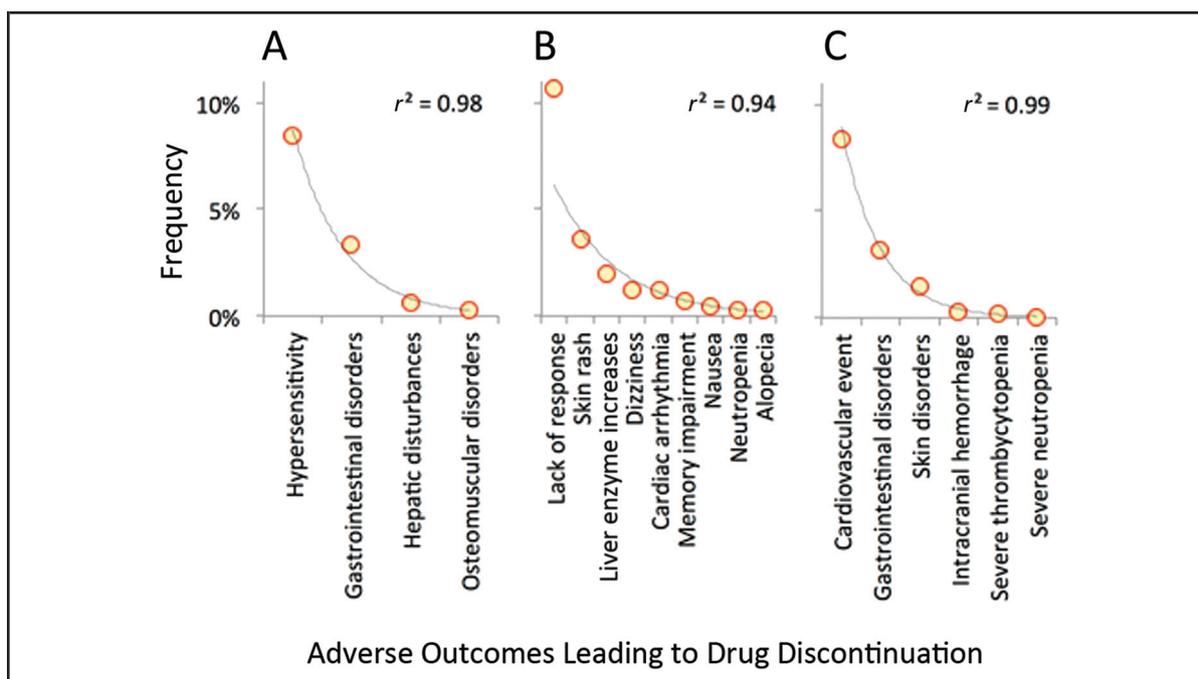
Fig. 1. Representative drug adverse-outcome profiles.

Shown are profiles for abacavir (A), carbamazepine (B), and clopidogrel (C). Exponential curves generally fit the profiles well ($r^2 = 0.88 - 0.99$). For example, the most common adverse outcome for clopidogrel (approximately 9%) is about 3 times more common ($e^{1.02}$) than the second most common adverse outcome (approximately 3%), which in turn is about 3 times more common than the third most common adverse outcome (approximately 1%), and so forth. Table S4 in the online Data Supplement lists all adverse outcomes and their frequencies (with references). See also Fig. S3 in the online Data Supplement.

we explored a large range of values in our simulations (30%–77%).

From PharmGKB *(34)* we estimated that candidate associations have already been discovered for adverse outcomes that account for 5% of all drug-related adverse outcomes (the first of our 4-step process). We took this as a starting point.

SIMULATION

We assigned adverse-outcome profiles to drugs (see Fig. S2 in the online Data Supplement) at random from the model (Fig. 1; see Fig. S3 and Table S4 in the online Data Supplement). For example, if a particular drug were assigned abacavir's adverse-outcome profile at random, that drug's most common adverse outcome would be modeled as affecting approximately 8% of its users, because abacavir's most common adverse outcome affects approximately 8% of abacavir users; its second-most common adverse outcome would be modeled as affecting approximately 3%, the same as for abacavir; and so on. We took this approach on the assumption that the adverse-outcome profiles of the drugs in our model are representative of drugs as a whole. This approach was an alternative to attempting

to compile profiles for every prescription drug. Summing over all drugs yielded the total incidence of adverse outcomes.

Next, for each adverse outcome we assigned a percent attributable risk due to a genetic association (Table 1). For example, for an adverse outcome that affects 10% of the users of a given drug, the percent attributable risk for a given genetic association might be taken to be 12% in the model (Fig. 2). This means that this genetic association explains 12% of the incidence of this adverse outcome; therefore, choosing or dosing of this drug according to the presence or absence of this variant would produce a 1.2% (i.e., 12% × 10%) decrease in adverse outcomes for this drug.

A cost and timeline for each step in the process of discovering and validating the association was assigned, again at random (see Table S2 in the online Data Supplement). These several steps were repeated, summing costs, until the total risk for this adverse outcome reached the maximum incidence that genetic guidelines were taken to be able to prevent (see above).

Sums for additional drugs and adverse outcomes were added to the total until the total incidence of adverse outcomes was accounted for (i.e., to model the
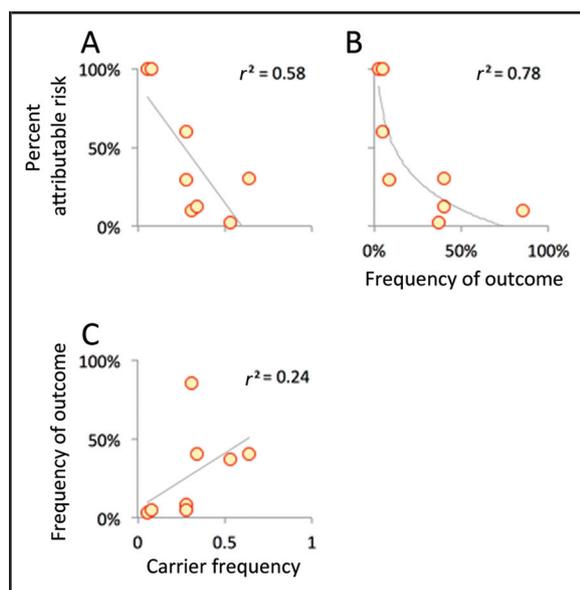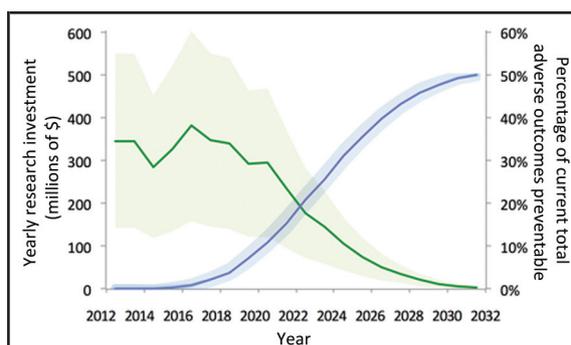
Fig. 2. Correlations between frequency of adverse outcomes, variant frequency, and percent attributable risk.

Negative correlations existed between percent attributable risk and variant frequency (A) and between percent attributable risk and adverse-outcome frequency (B); also noted was a weak positive correlation between adverse-outcome frequency and variant frequency (C). We accounted for these relationships in the model.

Fig. 3. Cost and benefit of developing pharmacogenomic drug treatment guidelines.

Monte Carlo methods were used to forecast the annual research investment for developing clinically validated pharmacogenomic guidelines (green, left axis) and the cumulative decrease in adverse outcomes to which they would lead when implemented (blue, right axis). The shaded area represents the overall 95% CI; both low- and high-cost scenarios are included. The line indicates the mean.

goal of halving the overall incidence of drug-related adverse outcomes). The result of this addition is the cost and timeline for the research required to halve drug-related adverse outcomes overall (Fig. 3). For statistical confidence, we repeated each simulation 1000 times (we note that means and SDs changed little beyond 100 repetitions). We confirmed that the labor force would not be a limiting factor for carrying out the required research and therefore would not affect the timeline (see Supplementary Methods and Fig. S4 in the online Data Supplement).

We performed sensitivity analysis by varying key parameters. Specifically, we varied the cost per coauthor (from $35,000 to $50,000) and the maximum incidence that genetic guidelines can prevent (for the goal of halving adverse outcomes, from the 60% of the base case to the 77% estimated for warfarin; this corresponds to 60/83 = 72% to 77/83 = 93% of the incidence of drug-related adverse outcomes that are currently considered nonpreventable) and by removal of each single profile in turn from the model and repeating the simulation. We used 2 SDs above and below the most- and least-expensive estimates, respectively, to establish overall 95% CIs for the model. We then simi-

larly explored a wider range of maximum incidence/percent attributable risk (30%–70%) for the more modest goal of cutting drug-related adverse outcomes by 25%.

Results

As a basis for our analysis we selected 8 drug–gene associations involving 7 adverse outcomes from the fields of cardiology, neurology, psychiatry, and infectious disease (Table 1). These included 5 of the 7 strongest associations identified by expert survey (32): *VKORC1*[4] (vitamin K epoxide reductase complex, subunit 1) and warfarin-induced bleeding; *CYP2C9* (cytochrome P450, family 2, subfamily C, polypeptide 9) and warfarin-induced bleeding; *HLA-B* (major histocompatibility complex, class I, B) and abacavir-induced hypersensitivity; *CYP2C19* (cytochrome P450, family 2, subfamily C, polypeptide 19) and cardiac events due to clopidogrel resistance; and *HLA-B* and carbamazepine-induced Stevens–Johnson syndrome. One association—between statin-induced myopathy and *SLCO1B1* (solute carrier organic anion transporter family, member 1B1)—involved several drugs (prava-

---

[4] Human genes: *VKORC1*, vitamin K epoxide reductase complex, subunit 1; *CYP2C9*, cytochrome P450, family 2, subfamily C, polypeptide 9; *HLA-B*, major histocompatibility complex, class I, B; *CYP2C19*, cytochrome P450, family 2, subfamily C, polypeptide 19; *SLCO1B1*, solute carrier organic anion transporter family, member 1B1; *COMT*, catechol-*O*-methyltransferase; *HTR2A*, 5-hydroxytryptamine (serotonin) receptor 2A, G protein-coupled.

statin, atorvastatin, and simvastatin) from a single class.

All adverse outcomes were defined by clinical rather than laboratory findings (e.g., warfarin-associated bleeding, as opposed to fraction of the time the international normalized ratio was outside the therapeutic range). Frequencies of adverse outcomes ranged from 2.7% (abacavir hypersensitivity) to 85% (return to smoking within 12 weeks of the nicotine-replacement patch). All associations involved a specific allele or alleles rather than copy-number variants or other types of genomic variation. Carrier frequencies ranged from 5.6% (*HLA-B*5701*) to 64% (*VKORC1*) in the studied populations. One association—between return to smoking after patch use and the *COMT* (catechol-*O*-methyltransferase) gene—was with a homozygous state, not a carrier state. The percent attributable risk due to the carrier state (or to the homozygous state for smoking and *COMT*) ranged from 2.3%—for the relationship of *HTR2A* [5-hydroxytryptamine (serotonin) receptor 2A, G protein-coupled] single-nucleotide polymorphisms rs2224721 and rs9316233 to the lack of response to escitalopram)—to 100% (*HLA-B*5701* in abacavir-induced hypersensitivity and *HLA-B*1502* in carbamazepine-induced hypersensitivity). There were inverse correlations between percent attributable risk and carrier frequency [$r^2 = 0.57$ for linear regression; bootstrap for single outliers $r^2 = 0.58$ (SD, 0.10)] and between percent attributable risk and adverse-outcome frequency [$r^2 = 0.71$ (SD, 0.07) for log-linear regression; Fig. 2]. Thus, more of the attributable risk was due to rare alleles than to common ones, especially for rarer adverse outcomes. We accounted for this in the model (see Supplementary Methods in the online Data Supplement).

A literature search yielded 29 reports that described the shortest necessary chains of evidence for discovering and validating each of these associations (see Supplementary Materials in the online Data Supplement). These reports covered a mean of 5225 patients per association (range, 1335–21 171). Clinical guidelines were reported for 6 of the 8 associations. Only 1 association—between abacavir-induced hypersensitivity and *HLA-B*5701*—had a guideline validated as useful for decreasing the incidence of the associated adverse outcome in clinical practice. Three other associations—*VKORC1* and warfarin-induced bleeding, *CYP2C9* and warfarin-induced bleeding, and *CYP2C19* and cardiac events due to clopidogrel resistance—have trials underway that could validate proposed guidelines (NCT01119300, NCT01006733, NCT01305148, and NCT00995514; see http://clinicaltrials.gov). Two of these trials closed in 2011.

| Table 2. Estimated cost of cutting drug-related adverse outcomes by 25%.[a]	
Adverse outcomes due to genomics, %	Total cost, billions of $
30	3.28 (0.36)
40	1.27 (0.17)
50	0.84 (0.13)
60	0.67 (0.14)
70	0.55 (0.11)
[a] See main text. Data are presented as the mean (SD).	

For the model, we assumed that reports validating these guidelines will appear before the end of 2012.

We found that the research investment necessary to develop guidelines capable of cutting drug-related adverse outcomes in half will require a research investment of $1.5 billion to $6 billion and take approximately 20 years (Fig. 3). This investment includes the cost spent and the time spent on unpublished and nonconfirmatory findings.

The single most important determinant of total cost was the overall percent attributable risk due to genomics: the extent to which and manner in which genomic variants are responsible for adverse outcomes. At the top of the range (77%), the total cost was approximately $2 billion; for a more conservative number (60%), the cost was $5 billion to $6 billion. The reason: the higher the percentage attributable risk, the larger the role genomics plays, and therefore the more high-impact associations there are likely to be, given the frequencies of the associations already found. Conversely, the smaller the role of genomics, the more rare associations will have to be found and confirmed to cut adverse outcomes in half, at a higher overall cost. Given this dependence, we also explored much lower percentage attributable risk in the context of the more modest, but still substantial, goal of cutting drug-related adverse outcomes by 25%. We again found costs in the range of single-digit billion dollars (Table 2).

By contrast, the total cost was much less dependent on the cost per coauthor and thus on the cost for producing each report. Bootstrapping by removing individual known associations from the model and repeating the simulation changed the dollar amount by <20% and the timeline by only about a year.

We projected that most of the research investment—$1.5 billion to $3 billion at $250 million to $500 million per year—will need to come well before most guidelines appear (Fig. 3). This is because the first 3 steps in the research process require time and have financial
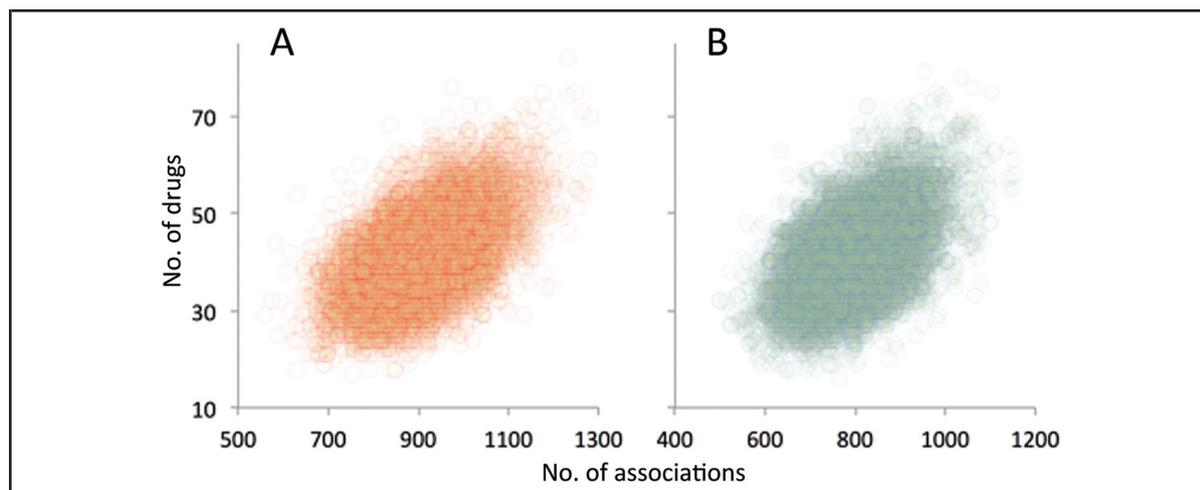
Fig. 4. Numbers of associations and drugs that guidelines must cover.

(A), Each point represents the total number of associations and drugs that pharmacogenomic guidelines will have to cover to be capable of halving the incidence of drug-associated adverse outcomes (assuming genetics can explain 60% of adverse outcomes). Each point is the result of a single simulation of the model. Results for 10 000 simulations are shown [mean (SD): for 910 (100) associations covering 41 (9) drugs, or a mean of 23 (4) associations per drug]. (B), Results were similar for the more modest goal of cutting adverse outcomes by 25%, on the assumption that genetics can explain 30% of adverse outcomes [mean of 802 (91) associations for 41 (9) drugs, for a mean of 20 (4) associations per drug; n = 10 000 simulations].

costs, but only completion of the fourth step leads to a clinically actionable guideline. Thus, by 2017 as much as $3 billion will have been invested as "pump priming," with only a few validated guidelines to show for it. Only in the ensuing 7 to 8 years will this investment pay off with the appearance of the bulk of the necessary guidelines, making possible a substantial reduction in adverse outcomes between 2018 and 2025.

We found that the number of researchers in pharmacogenomics and related areas has grown exponentially since 1980 ($r^2 = 0.95–0.96$; see Fig. S4 in the online Data Supplement) resulting in an estimated 15 424–30 584 researchers by the end of 2012. At a mean of 13.4 authors per report (see Fig. S1 and Supplementary Methods in the online Data Supplement), the labor force is sufficient to produce the needed guidelines without requiring further growth in the number of researchers. Thus, the labor force does not need to continue to grow at the observed rate to achieve the stated goal in the stated time; it simply must not shrink.

Discussion

Quantitative modeling has become a standard tool for planning and for setting expectations in many fields, from politics to meteorology. Here we apply it to genomics—specifically pharmacogenomics—to estimate the time and financial investment required to use

genomics to substantially reduce the rate of drug-related adverse outcomes. These adverse outcomes currently cost the healthcare system tens of billions of dollars each year and are potentially low-lying fruit for genomic medicine. After testing a range of situations, we estimate the total cost will be in the single-digit billions of dollars, front-loaded but spread out over approximately 20 years. Specifically, we expect a total cost of ≤$6 billion. This cost includes the cost of failures and unpublished work but excludes the work of implementing guidelines and the (rapidly falling) costs of sequencing.

The single most important factor influencing cost in our model is the extent to which genetics contributes to drug-related adverse outcomes, measured as the percentage of drug-related adverse outcomes that can be traceable to genomic variants. Research suggests that 60% is a reasonable estimate. We found that as the percentage falls even modestly, the cost increases sharply. If genomics' contribution is <50% on average, halving adverse outcomes through genomics is not feasible. However, our model also lets us estimate the cost required for smaller-scale goals, such as for example a 25% reduction in drug-related adverse outcomes (approximately $0.5 billion to $3 billion; Table 2). The major assumption is that the measurements of drugs and associations investigated thus far are representative of pharmacogenomics as a whole. Better estimates of the contribution of genomics would be useful.

We find the most important bottleneck is the discovery and confirmation of candidate associations between genetic variants and adverse outcomes. To date, this task has been a slow, iterative process. In principle the process could be accelerated by studying several thousand people taking each of the 40–50 (Fig. 4) most-used prescription drugs (epidemiologic and statistical studies would be needed to determine the precise number of drugs and the number of patients necessary to ensure representation of the most common adverse outcomes across ethnicities). Because the time scale of many adverse outcomes is only days to months, the time scale for such a project would be set by the time required to raise funds, to build a consortium, to identify and register participants, and to collect, analyze, and interpret samples. A "50 000 Pharmacogenomes Project" would be a nontrivial undertaking. But in the spirit of the 1000 Genomes Project, UK10K (35), or the Million Veteran Program, it would have the potential to represent an improvement (36) that would disrupt today's process (which we have modeled) and thereby save time and money.

Numbers invite comparisons. Six billion dollars represents an annualized 4% of the NIH's 2011 budget (37), 2% of the pharmaceutical industry's 2011 research budget (38), and 0.2% of the 2009 payments made by US private health insurers (39). Our 20-year timeline is consistent with qualitative forecasts by the National Human Genome Research Institute that genomics will produce some advances in the science of medicine before 2020 but many more thereafter (6). Finally, $6 billion over 20 years is comparable to the cost and timeline for developing just 4 to 5 new drugs from scratch (40). Whether the benefit of developing guidelines is worth $1.5–$6 billion, if so, who should pay (government, insurers, hospitals, industry) and if not, what process improvements would change that are open questions.

Genomics is maturing. As the fruits of research enter clinical care, predictions become easier. Our model is not the last word but we hope, the start of a conversation and an illustration of the value of quantitative modeling in genomics. Modeling for cancer genomics, metagenomics, and heritable conditions will soon be possible. Such models should help decision makers and the public set expectations and priorities for translating genomic research into better patient care.

References

1. Wade N. A decade later, genetic map yields few new cures. New York Times 2010 June 13; Sect. A:1.
2. Pollack A. Awaiting the genome payoff. New York Times 2010 June 15; Sect. B:1.
3. The genome, 10 years later [Editorial]. New York Times 2010 June 21; Sect. A:28.
4. Evans JP, Meslin EM, Marteau TM, Caulfield T. Genomics. Deflating the genomic bubble. Science 2011;331:861–2.
5. Hall SS. Revolution postponed. Sci Am 2010;303: 60–7.
6. Green ED, Guyer MS. Charting a course for genomic medicine from base pairs to bedside. Nature 2011;470:204–13.
7. Boguski MS, Arnaout R, Hill C. Customized care 2020: how medical sequencing and network biology will enable personalized medicine. F1000 Biol Rep 2009;1:73.
8. Pfanner E. The rise of the quants in political prognostication. IHT Rendezvous (International Herald Tribune: the global edition of the New York Times. November 8, 2012. http://rendezvous.blogs.nytimes.com/2012/11/08/the-rise-of-the-quants-in-political-prognostication/ (Accessed December 2012).
9. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH. Big data: the next frontier for innovation, competition, and productivity. McKinsey Global Institute; 2011. http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation (Accessed December 2012).
10. van der Duin P. Knowing tomorrow? How science deals with the future. Delft: Eburon Publishers; 2009.
11. Porter AL, Ashton WB, Clar G, Coates JF, Cuhls K, Cunningham SW, et al. Technology futures analysis: toward integration of the field and new methods. Technol Forecast Soc Change 2004;71: 287–303.
12. Ginsburg GS, Willard HF. Genomic and personalized medicine: foundations and applications. Transl Res 2009;154:277–87.
13. Roden DM, Johnson JA, Kimmel SE, Krauss RM, Medina MW, Shuldiner A, Wilke RA. Cardiovascular pharmacogenomics. Circ Res 2011;109: 807–20.
14. Al-Ghoul M, Valdes R Jr. Fundamentals of pharmacology and applications in pharmacogenetics. Clin Lab Med 2008;28:485–97.
15. Kasotakis G, Pipinos II, Lynch TG. Current evidence and clinical implications of aspirin resistance. J Vasc Surg 2009;50:1500–10.
16. Verschuren JJ, Trompet S, Wessels JA, Guchelaar HJ, de Maat MP, Simoons ML, Jukema JW. A systematic review on pharmacogenetics in cardiovascular disease: Is it ready for clinical application? Eur Heart J 2012;33165–75.
17. LaRosa JC, Grundy SM, Waters DD, Shear C, Barter P, Fruchart JC, et al. Intensive lipid lowering with atorvastatin in patients with stable coronary disease. N Engl J Med 2005;352:1425–35.
18. Bates DW, Cullen DJ, Laird N, Petersen LA, Small SD, Servi D, et al. Incidence of adverse drug events and potential adverse drug events. Impli-

cations for prevention. ADE Prevention Study Group. JAMA 1995;274:29–34.

19. Tache SV, Sonnichsen A, Ashcroft DM. Prevalence of adverse drug events in ambulatory care: a systematic review. Ann Pharmacother 2011;45:977–89.

20. Bates DW, Spell N, Cullen DJ, Burdick E, Laird N, Petersen LA, et al. The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. JAMA 1997;277:307–11.

21. Feero WG, Guttmacher AE, Collins FS. Genomic medicine—an updated primer. N Engl J Med 2010;362:2001–11.

22. Mrazek DA, Lerman C. Facilitating clinical implementation of pharmacogenomics. JAMA 2011;306:304–5.

23. Paugh SW, Stocco G, McCorkle JR, Diouf B, Crews KR, Evans WE. Cancer pharmacogenomics. Clin Pharmacol Ther 2011;90:461–6.

24. Wilke RA, Dolan ME. Genetics and variable drug response. JAMA 2011;306:306–7.

25. Holmes MV, Casas JP, Hingorani AD. Putting genomics into practice. BMJ 2011;343:d4953.

26. Wadelius M, Chen LY, Eriksson N, Bumpstead S, Ghori J, Wadelius C, et al. Association of warfarin dose with genes involved in its action and metabolism. Hum Genet 2007;121:23–34.

27. Wang L, McLeod HL, Weinshilboum RM. Genomics and drug response. N Engl J Med 2011;364:1144–53.

28. Gullov AL, Koefoed BG, Petersen P. Bleeding during warfarin and aspirin therapy in patients with atrial fibrillation: the AFASAK 2 study. Atrial Fibrillation Aspirin and Anticoagulation. Arch Intern Med 1999;159:1322–8.

29. Smith P, Arnesen H, Holme I. The effect of warfarin on mortality and reinfarction after myocardial infarction. N Engl J Med 1990;323:147–52.

30. Mallal S, Phillips E, Carosi G, Molina JM, Workman C, Tomazic J, et al. HLA-B*5701 screening for hypersensitivity to abacavir. N Engl J Med 2008;358:568–79.

31. Burke W, Laberge AM, Press N. Debating clinical utility. Public Health Genomics 2010;13:215–23.

32. Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. Clin Pharmacol Ther 2011;89:464–7.

33. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet 2001;29:306–9.

34. Klein TE, Chang JT, Cho MK, Easton KL, Fergerson R, Hewett M, et al. Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base. Pharmacogenomics J 2001;1:167–70.

35. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56–65.

36. Topol EJ. The creative destruction of medicine: How the digital revolution will create better health care. New York: Basic Books; 2012.

37. HHS.gov. U.S. Department of Health & Human Services. FY2012 president's budget for HHS: HHS budget in brief. http://www.hhs.gov/about/budget/index.html#HHSBudgetinBriefandPerformanceHighlights (Accessed December 2012).

38. Pharmaceutical Research and Manufacturers of America (PhRMA). About PhRMA. http://www.phrma.org/about/about-phrma (Accessed March 2012).

39. CMS.gov: Centers for Medicare & Medicaid Services. National health expenditure data. https://www.cms.gov/NationalHealthExpendData/25_NHE_Fact_Sheet.asp (Accessed March 2012).

40. Vernon JA, Golec JH, Dimasi JA. Drug development costs when financial risk is measured using the Fama-French three-factor model. Health Econ 2010;19:1002–5.

41. Hulot JS, Collet JP, Silvain J, Pena A, Bellemain-Appaix A, Barthelemy O, et al. Cardiovascular risk in clopidogrel-treated patients according to cytochrome P450 2C19*2 loss-of-function allele or proton pump inhibitor coadministration: a systematic meta-analysis. J Am Coll Cardiol 2010;56:134–43.

42. Link E, Parish S, Armitage J, Bowman L, Heath S, Matsuda F, et al. SLCO1B1 variants and statin-induced myopathy—a genomewide study. N Engl J Med 2008;359:789–99.

43. Uher R, Huezo-Diaz P, Perroud N, Smith R, Rietschel M, Mors O, et al. Genetic predictors of response to antidepressants in the GENDEP project. Pharmacogenomics J 2009;9:225–33.

44. Johnstone EC, Elliot KM, David SP, Murphy MF, Walton RT, Munafo MR. Association of COMT Val108/158Met genotype with smoking cessation in a nicotine replacement therapy randomized trial. Cancer Epidemiol Biomarkers Prev 2007;16:1065–9.

45. Munafo MR, Johnstone EC, Guo B, Murphy MF, Aveyard P. Association of COMT Val108/158Met genotype with smoking cessation. Pharmacogenet Genomics 2008;18:121–8.

46. Wadelius M, Chen LY, Lindh JD, Eriksson N, Ghori MJ, Bumpstead S, et al. The largest prospective warfarin-treated cohort supports genetic forecasting. Blood 2009;113:784–92.

47. Kangelaris KN, Bent S, Nussbaum RL, Garcia DA, Tice JA. Genetic testing before anticoagulation? A systematic review of pharmacogenetic dosing of warfarin. J Gen Intern Med 2009;24:656–64.

48. Aithal GP, Day CP, Kesteven PJ, Daly AK. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. Lancet 1999;353:717–9.

49. Chen P, Lin JJ, Lu CS, Ong CT, Hsieh PF, Yang CC, et al. Carbamazepine-induced toxic effects and HLA-B*1502 screening in Taiwan. N Engl J Med 2011;364:1126–33.