A randomised survey of the quality of antibiotics and other essential medicines in Indonesia, with sales volume-adjusted estimates of the prevalence of substandard medicines.

Elizabeth Pisani, Ayu Rahmawati, Esti Mulatsari, Mawaddati Rahmi, William Nathanial, Yusi Anggriani, on behalf of the STARmeds Study Group.

Author affiliations: All named authors are from the Faculty of Pharmacy, Universitas Pancasila, Jakarta, Indonesia. Additional members of the STARmeds Study Group are listed, by affiliation, at the end of the manuscript

Abstract

Introduction

Most substandard and falsified medicines are at best not optimally effective, and at worst fatal. While the World Health Organisation and others warn they are a major threat to public health in low and middle income countries, little is known about their true prevalence. Authors of meta-analyses universally warn that survey data are not generalisable, because of unrepresentative study designs and variations in medicines included; tests performed; reference standards and pharmacopeia used; and definitions used when translating multiple quality parameters into a single pass/fail measure.

We hypothesised that weighting for sales volume of different products and brands would increase accuracy of estimates of medicine quality.

Methods and Findings

We collected samples of allopurinol, amlodipine, cefixime and dexamethasone, as well as amoxicillin in 2 formulations in seven districts across Indonesia, the world's fourth most populous country. Outlets, including retail pharmacies, over the counter medicine shops, public and private hospitals, primary health centres, doctors and nurses were randomised. We also sampled from the internet. Retail samples were collected by mystery shoppers, other samples overtly.

We tested 1274 samples for identity and assay, and all relevant samples for dissolution and uniformity of content, using USP reference standards and monographs. Samples that failed any laboratory test were considered out of specification. We calculated prevalence per product and brand, and weighted the results by the sales volume of each product, using sales data from IQVIA and the national public procurement system.

The weighted prevalence of out-of-specification products was 4.9%, 40.3% lower than the raw estimate (8.2%). Antibiotics were more likely to be substandard (weighted prevalence 6.8 vs 3.1; raw prevalence 13.6 vs 4.9, $p<0.000$). There was no relationship between quality and any of the following: price; branded status; public procurement status; outlet type. Our estimate compared with the Indonesian medicine regulator's estimate of 4.0%, calculated based on unweighted analysis of 13,539 samples of a wider variety of medicines, collected overtly nation-wide.

Conclusions

Where data are available, weighting survey results by sales volume is a cost-effective way of improving the accuracy of estimates of out-of-specification medicines measured in field surveys.

A randomised survey of the quality of antibiotics and other essential medicines in Indonesia, with sales volume-adjusted estimates of the prevalence of substandard medicines.

Introduction

In 2017, the World Health Organization (WHO) estimated that 10.5% of medicines in all low and middle income countries were substandard (they did not meet the standards laid out in their market authorisation paperwork) or falsified (they deliberately misrepresented content, identity or source).[1] That implies that out of every 10 pills a patient swallows, every 10 spoonfuls of medicinal syrup a child drinks, every 10 injections a nurse administers, one will be at best not optimally effective, and at worst life-threatening. As governments in low and middle income countries strive to expand access to medicines for 7.3 billion citizens, this figure, if indeed correct, should surely be a cause for significant concern, as well as increased investment in quality assurance.

The 2017 estimate was based on a review of 100 papers published between 2007 and 2016, reporting studies that included between 10 and 3,024 tested samples, with a median size of 123 samples (Q1–Q3: 30-325). While substandard and falsified medicines are occasionally identified in higher income settings, often in on-line markets, it is widely assumed that robust regulatory oversight effectively secures the quality of the overwhelming majority of medicines in the regulated supply chain in those countries. Low and middle income countries are assumed to have lower regulatory capacity, hence be less able to ensure quality in production, or provide oversight for imported medicines and throughout the supply chain.[2] In the studies collated by WHO, 47.1% of samples were collected in sub-Saharan African countries, home to around 15% of the population of all low and middle income countries. Over 71% of 34,300 tested samples with stated active ingredients were for three of the diseases of greatest interest to large global health funders: HIV, TB and malaria. Fewer than half of the included studies attempted randomisation of outlets. Most studies sampled from the retail sector only; none included sampling from the internet.

All studies reported raw prevalence: the number of samples failing any test conducted in the study, divided by the number tested (a sample being a sample of a single active ingredient, dose, formulation and brand/market authorisation holder, collected at one time and place). None adjusted for the repeat collection of different samples of the same product, and none were reported to take into account the market share of a particular brand when calculating results. Data extracted from these studies, including the number of samples tested and failing testing, broken down by therapeutic category, country, and broad sampling and testing methods, is available in our supporting materials archive: https://doi.org/10.7910/DVN/QRKDWG.

More recent reviews that draw on many of the same studies, and include those focusing on specific classes of medicines, have found similar results (with overall aggregate prevalences of poor quality antibiotic, cardiovascular and diabetes medicines in low and middle income countries of 17.4, 15.4, and 10.8% respectively). [3–6] A review of the quality of antiretrovirals (many of which are procured for vertical, donor-funded programmes from WHO pre-qualified manufacturers), found a prevalence of 1.4% out of specification.[7] All the cited reviews point out that the figures published in academic medicine quality surveys are unlikely to be representative or generalisable to settings beyond the specific study setting. Reasons for this include small sample sizes, itself a

consequence of the high costs of pharmacopeial testing, as well as unrepresentative study designs and variations in medicines included, tests performed, reference standards and pharmacopeia used and definitions used when translating multiple quality parameters into a single pass/fail measure.

Overall quality of studies, including the proportion selecting outlets randomly, has increased over time.[8] However there are still major gaps in our understanding of the true level of the threat posed by medicines that are substandard or falsified. From a regulatory point of view, it may not be necessary to have a precise estimate of the prevalence of poor quality medicines in a market. Indeed, many well-established regulators, including the US Food and Drug Administration, make no attempt to implement active post-market surveillance by collecting and independently testing samples of medicines in circulation in the supply chain, nor do they develop national estimates of prevalence. Instead, they rely on industry and health care providers to report information about the performance of medicines in the market after approval.[9] Other regulators choose to operate active case-finding systems; instead of testing products randomly sampled from the supply chain, they use risk algorithms to actively seek out products most likely to be substandard or falsified, in order to remove them from the supply chain. [10,11]

However, global health actors and researchers continue to call for survey data that would provide a more nuanced understanding of the actual prevalence and distribution of poor quality medicines circulating in specific markets, reasoning that such data could underpin advocacy for better regulation, and the resources it requires.[12,13] Some national governments also regard such estimates as an important metric in evaluating the performance of the medicine regulator.[14] But the cost of testing medicines, combined with methodological challenges and sometimes poor information on the volumes of medicines in circulation, undermine regulators' and researchers' ability to test a truly representative sample of medicines. Unless prevalence is high, large sample sizes are needed for robust estimates. Randomisation at the level of the medicine within each retail outlet is generally incompatible with the collection of samples by mystery shoppers, a method recommended by WHO and scholars.[15,16] It is not always feasible to sample from all the outlets where patients acquire medicines, including formal and informal retail outlets, both physical and online, and public and private health facilities.

In addition, surveys typically try to sample a range of different brands. However, the market share of each brand may be very different, with implications for interpretation of survey data. For example, a poor quality product that was distributed free to all patients in the public health system nationwide would likely cause greater harm than a similar quality failure in a high-end brand provided only to a small number of rich patients in the capital city.

At least two studies have attempted to account for these variations in market share, the first, in DRC Congo, by weighting prevalence by distributor sales volume,[17] and the second, in Indonesia, in its sampling design[18]. In Congo, researchers found that poor quality products had smaller distribution, so the overall prevalence of substandard medicines was lower than raw prevalence estimates would suggest. Authors of the Indonesian study were unable to complete planned adjusted estimates of the prevalence of poor quality cardiovascular and diabetes medicines because all 204 samples of 5 medicines passed both assay and dissolution testing. At the same time, several scandals related to medicine falsification in Indonesia, and concerns about very low prices resulting from public procurement practices described at greater length below, were

causing public concern about the quality of medicines across Indonesia, the world's fourth most populous nation.

In light of this uncertainty, and the larger question around the representativeness of medicine quality studies, we conducted a large survey in Indonesia, testing 1,274 samples of five medicines, including antibiotics, steroids, and treatment for non-communicable diseases. We collected samples in urban and rural areas across the country from all types of outlets from which patients acquire medicines. We adjusted our raw prevalence data for the market distribution of different brands nationally to better reflect true prevalence, and investigated the effect of using different definitions of product quality. We refer to the study as STARmeds (for Systematic Tracking of At-Risk medicines).

Methods

All methods are described in greater detail in Supplementary Document 1, according to MEDQUARG guidelines. The supplementary file provides details of secondary data sources, medicine and site selection, sample size calculation, sample frame construction, sample collection and handling, data entry and management, laboratory testing, ethics protections and reporting procedures, as well as estimation methods. We summarise these briefly here. Additional documentation, including a MEDQUARG reporting checklist, data cleaning and management code, estimation code, laboratory protocols and sample-level data (with brands masked in accordance with the requirements of our ethics approval) is provided in the study archive, at https://dataverse.harvard.edu/dataverse/STARmeds.[19]

In addition, we have published a toolkit providing detailed, practical guidance on conducting medicine quality surveys. This includes generic, freely downloadable and adaptable versions of all of our study tools (data collection software, data management code, sample control forms etc.).[20]

Study setting

Indonesia is home to one of the world's largest and most generous public health insurance schemes; it also has a vibrant domestic pharmaceutical sector. National health insurance (*Jaminan Kesehatan Nasional* or JKN), was introduced in 2014, and by the end of 2022 covered 249 million people, 90% of the population.[21] During that period, national tenders and procurement through a single-winner electronic platform pushed down the price of most medicines, while increasing volumes consumed.[22] In 2022, Indonesian health service providers bought a total of 7.6 billion units of medicine through the platform, at a total cost of US$519 million. Domestically produced medicines accounted for 99% of the products by volume, but only 56% by value. Because many patients buy medicines outside of the public system, and because unit prices in the single-winner tender system were very low, the public procurement sales remained a small fraction of the total market, estimated at US$3.6 billion in 2022.[23]

The quality of both domestically produced and imported products are overseen by Indonesia's medicine regulator (*Badan Pengawas Obat dan Makanan*, or BPOM), BPOM inspects factories, and conducts extensive post-market surveillance, collecting and testing medicines in circulation in the supply chain using both random and risk-based sampling. The random sampling -- accounting for 10,980 medicine samples in 2021 (excluding traditional medicines and dietary supplements), stratifies down to the

district level, then samples both outlets and medicines randomly. 3.7%of these samples did not meet all standards in 2021 (a measure that included not having a valid registration number, being expired at the time of sampling, being physically damaged, having errors in labelling or failing at least one pharmacopeial test). Another 2,559 samples were collected using risk-based sampling, which focuses on products of public health importance at high risk for being of poorer quality (including products from less experienced manufacturers, or those with previous failures); in this group, 5.2% failed, for an overall estimate of 4.0% of products nationally that did not meet regulatory standards.

The fall in prices for domestically-produced medicines that followed the introduction of JKN led manufacturers to warn that they may not be able to sustain quality production.[24–26] Together with a falsified vaccine scandal in 2017, these warnings undermined public confidence in the regulator.[27] Then, in 2022, a rise in cases of acute kidney injury in children was traced to contaminated medicines given to children to treat coughs and fevers. Over 320 cases were recorded, at least 200 people, mostly children, died, underlining the potentially dire consequences of substandard medicines.[28] Similarly contaminated paediatric syrups products identified in Iraq, Cameroon, Gambia, Uzbekistan, Cambodia and the Marshall Islands in 2022/23 related mostly to products made in India, but all the Indonesian deaths were associated with medicines made by licensed domestic producers.[29] Once alerted to the deaths by the Ministry of Health, the Indonesian regulator acted quickly to identify and shut down the source of contaminated products.[30] However the case raised the spectre of a poorly-regulated market awash in dangerous products, leading to a class action suit against the medicine regulator.[31,32]

The perception of high rates of substandard medicines seemed to be at odds with the findings of a recent independent field survey of cardiovascular and diabetes medicine quality in Indonesia, which found no substandard products at all,[18] as well as with those of BPOM's annual post-market surveillance, which recorded failure rates of 4%. However, BPOM only samples from the regulated supply chain, excluding informal retail and wholesale markets, the internet, and medicines supplied to patients by doctors, nurses and midwives; the representativeness of their post-market surveillance system has been called into question.[14,33]

Study design

Study medicines were chosen based on public health importance, diversity of suppliers and feasibility, as described in the Supplementary methods. We also considered risk of falsification. Included products are shown in Table 1; all require prescriptions in Indonesia. When selecting medicines and sampling locations, we consulted with the Indonesian medicine regulator BPOM, and an intersectoral working group on medicine quality.

Table 1: Products included in the STARmeds study

Active ingredient	Primary use	Target doses	# of registered products*	Falsification risk
Allopurinol	Anti-hyperuricemia (Gout)	100mg tablet; 300mg tablet	65 46	Used non-medically
Amlodipine	Anti-hypertensive	5mg tablet	112	None
Amoxicillin	Antibiotic	500mg tablet/capsule; 125mg dry syrup	85 68	None
Dexamethasone	Anti-inflammatory	0.5mg tablet	59	None
Cefixime	Antibiotic	100mg tablet/capsule	45	Not free at primary level; relatively expensive

*Number of different brands/branded generics of the target doses and formulations registered in the Indonesian market, from the public domain BPOM product registration database (2022)

Three products had no special risk for falsification. Allopurinol is a cheap treatment for gout which was at the time of research being used off-label as a sedative in some urban areas in Indonesia. Cefixime is a relatively expensive antibiotic that was reported to be increasingly frequently prescribed to patients, but was not covered by insurance at the primary level, and was widely searched for on-line (see supplementary methods for more detail).

Sampling locations were chosen purposively to reflect Indonesia's geographic and economic diversity. Within each selected area, we chose outlets randomly, as discussed below. Table 2 shows locations, characteristics, sampling dates and type of outlet randomisation.

Table 2: Sampling locations, characteristics and sampling dates

District/ sampling area	Geographic area	Population per km/sq.*	Annual per capita GDP (US$)**	Outlet randomisation method	Sampling dates, 2022
Greater Jakarta	Central megacity	14,792	7,784	Two-stage PPS	15-20 February
Surabaya city	Large city	8,225	15,270	SRS	1-5 March
Malang regency	Semi-rural	733	2,941	One stage PPS	1-5 March
Medan city	Large city	8,525	7,553	SRS	22-26 March
Labuhan Batu regency	Remote rural	225	5,534	Take all for pharmacies and SRS for other outlets	22-26 March
Kupang City	Small city	2,335	3,784	SRS	29 March-2 April
Timor Tengah Selatan regency	Remote rural	118	1,289	Take all for pharmacies and SRS for other outlets	5-8 April

*2022 data from BPS/StatisticsIndonesia
** 2022 data from BPS/StatisticsIndonesia Rupiah values converted at Bank Indonesia average rate for 2022: 1 USD = 14,870.61 rupiah; Greater Jakarta is weighted average for sampling districts.
PPS: Probability proportionate to size. SRS: Simple random sampling.

Sample frame construction

Our sample was randomised at the level of the outlet.

We aimed to sample from all outlets from which Indonesian patients commonly acquire medicines, whether or not they are technically permitted to dispense prescription medicines. In each of the selected districts (or subdistricts, if using two-stage random sampling), we listed and verified retail outlets (pharmacies, over-the-counter medicine

shops, individual shops in bulk medicine markets); public and private hospitals; primary health centres; and doctors and midwives.

We additionally sampled from the internet and apps. We included all registered apps offering instant delivery of medicines using geo-positioning where available; all other registered online pharmacies; and all unlicensed platforms selling medicines identified through Google searches on terms such as "buy medicine", "buy antibiotic" "medicine for x". Sellers on general internet marketplaces included the on-line stores of licensed pharmacies. With very few exceptions, these are not technically licensed to sell online by Indonesian authorities. However, Tokopedia -- the biggest online sales platform in Indonesia -- operates its own verification system for pharmacies. Unverified individual sellers also operate on these platforms. For internet sources, we bought (or attempted to buy) samples from every one of the identified selling sites, but not from all individual sellers on those sites.

We distributed the target number of samples per medicine across the randomly selected retail and online outlets in clinically plausible combinations. We aimed to sample medicines at two price points (more and less expensive) from each retail outlets; two different buyers visited each selected outlet on different days. In health facilities, sampling was overt and we aimed to collect a branded and an unbranded version of every study medicine where available.

Of the outlets sampled, over-the-counter medicine shops and stores in bulk medicine markets that are not fully licensed pharmacies, as well as doctors, midwives and all but seven of the online sellers were not technically permitted to dispense prescription medicines at the time of sampling. The medicine regulator BPOM does not sample from these outlets in routine post-market surveillance.

Sample collection and handling

We define a sample as a single product (active ingredient, dose, formulation and brand or, if unbranded, market authorisation holder), collected in one place at one time. Since patients do not commonly check batch numbers when buying medicines, it was not always feasible to ensure that all tablets, capsules or bottles of syrup in a single sample were of the same batch. The ideal and minimum number of tablets/capsules/bottles to be purchased per sample are provided in the detailed methods annex.

Mystery shoppers native to the study area were trained online, and then in a full-day face-to-face session including role-play and hands-on practise of all study procedures. Each was assigned a single price-point (either cheap or expensive) for the duration of sampling. They were instructed to dress appropriately for the target price point, and given a daily personalised sample frame. They entered a selected shop and requested medicines using pre-prepared scenarios, including buying medicines for a sick relative, or stocking up for a journey. They signalled desired price points using phrases such as "Is this the very best brand you have?" or "Do you have anything more affordable?" Shoppers carried prescriptions for the target medicines, but did not offer them unless requested by the sales staff.

After exiting the shop, they put each sample (all strips, blisters or bottles of each unique medicine) in a separate Ziploc bag pre-labelled with a barcode. They scanned the barcode using the smart-phone based Open Data Kit/KoboCollect software,[34] geolocated the pharmacy, and entered medicine and price details. All Ziplocs were

delivered to a study hub in each sampling district in person or by courier at the end of each day.

Study staff collected samples overtly from health facilities. We submitted formal request letters to the directors of hospitals and health centres. Doctors and midwives were approached by phone from public listings if possible; otherwise study staff visited their surgeries and explained the purpose of the study. We paid for all samples, unless the facility preferred that we provide replacement product. Samples were then processed in the same way as retail samples.

Other than those bought through geo-positioned apps, most online samples were collected by study staff based in Jakarta. They used a separate KoboCollect form to record details of chats with sellers, orders, prices, and shipment costs. Purchases were sent to a variety of addresses, before being delivered to the Jakarta study hub.

At each district hub, the site supervisor inspected each incoming sample visually for anomalies in packaging or labelling, using a magnifying glass as necessary. No reference packaging from manufacturers was available for comparison. Dedicated data entry staff entered further details about each sample, including batch numbers, expiry dates and maximum retail price, using tablets pre-loaded with the study software in KoboCollect. From within the app, they photographed primary packaging, batch numbers and expiry dates in high resolution, using a portable light-box. Another staff member verified each sample before storing in a container containing a temperature logger, in an airconditioned room.

Samples from outside Greater Jakarta were hand-carried or air-freighted to Jakarta for further triage and transfer to the laboratory. For budgetary reasons, we excluded some samples from laboratory testing, using a systematic triage system described in the supplementary methods.

Laboratory testing

Samples were tested using United States Pharmacopeia (USP) reference standards, at PT Equilab International, a private laboratory in Jakarta which is ISO/IEC 17025: 2017 certified (and is also WHO prequalified for bioavailability testing). Laboratory staff inspected medicines visually for integrity and markings before preparing samples for pharmacopeial testing. All except for cefixime capsules were tested according to USP 43, NF38 monographs.[35] No USP monograph is available for cefixime capsules. For these, we followed Supplement 1 of *Farmakope Indonesia* 6th edition that refers to the Chinese Pharmacopeia.[36] The acceptance criteria for each product are shown in Table 3.

Identification and assay (including uniformity of content) were tested using high performance liquid chromatography (HPLC-UV Waters, Alliance 2695 with UV Detector 2489). HPLC was also used in dissolution testing for amoxicillin tablets and dexamethasone. Dissolution of amoxicillin capsule, allopurinol, amlodipine and cefixime was analysed using Spectrophotometer-UV/VIS (Shimadzu UV-1800). No dissolution testing was performed on amoxicillin dry syrup formulation. Details of equipment and full operating procedures are provided in supplementary methods and the study archive.

We could not afford to test for impurities.

Table 3: Acceptability criteria for pharmacopeial tests, USP 43 NF38

Active ingredient	Identification	Assay (%)	Dissolution (%) ('Q')	Content Uniformity
Allopurinol	Retention time of the major peak of the sample solution corresponds to that of the reference solution	93.0 – 107.0	75% in 45 minutes	NA
Amoxicillin, tablet		90.0 - 120.0	75% in 30 minutes	NA
Amoxicillin, capsule		90.0 - 120.0	80% in 60 minutes	NA
Amoxicillin, dry syrup		90.0 - 120.0	NA	NA
Cefixime, tablet		90.0 – 110.0	75% in 45 minutes	NA
Cefixime, capsule		90.0 – 110.0	80% in 45 minutes	NA
Amlodipine		90.0 – 110.0	75% in 30 minutes	Acceptance value ≤ 15.0, and no individual tablet has an assay value that falls outside USP-specified limits.
Dexamethasone		90.0 - 110.0	80% in 30 minutes	

NA: Not applicable

Data handling

Data from the field, research hub and lab apps were downloaded in comma separated value format, imported into Stata 17,[37] merged on sample barcode, and cleaned and coded as necessary.

We shared high resolution images of primary packaging, as well as information on batch number, expiry date and maximum retail price for every sample with all 79 market authorisation holders, and requested that they verify that the data accord with their production records.

Analysis and estimation

Price variation

To enable the comparison of prices across medicines with different base prices, we calculated the ratio of the price paid for each sample to the median price for all samples of that medicine, dose and formulation (tablet or dry syrup), including untested samples. In calculating ratios, we replaced the price of medicines provided free to insured patients with the price paid in public procurement in the province of sampling plus 28%, reflecting the addition of tax and dispensing charges which regulations allow for retail medicines. We made no distinction between packaging (blister or foil), or tablet verses capsule or caplet formats.

Product quality

Our core definition for an out-of-specification sample was a sample that failed any laboratory test to which it was submitted, using the limits shown in Table 3.

Raw prevalence was calculated by dividing the number of substandard samples by the number tested. For samples that failed any pharmacopeial test, we also calculated the magnitude of deviation from permitted values by choosing the greatest deviation among any of the following values:

- percentage points by which assay (the percent of labelled active ingredient identified in the sample) exceeds the upper limit of acceptability;
- percentage points by which assay falls short of the lower limit of acceptability;

- percentage points by which final dissolution value (the % of labelled active ingredient dissolved within the allotted testing time) falls below the dissolution threshold;
- points by which uniformity acceptance value (a measure of the acceptable variation between pills in a single sample in the % of labelled active ingredient identified) exceeds 15.

To examine the effect of including additional quality parameters, we also report prevalence by two additional measures, which we refer to as Expanded, and Maximum. The parameters included in each of the three measures are summarised in Table 4.

Table 4: Definitions of quality use in analysis

	Measure of quality	Core	Expanded	Maximum
	Failed assay testing	Yes	Yes	Yes
OR	Failed dissolution testing	Yes	Yes	Yes
OR	Failed uniformity of content testing	Yes	Yes	Yes
OR	Expired at time of purchase,	No	Yes	Yes
OR	No expiry date on primary packaging	No	Yes	Yes
OR	Broken or damaged at time of purchase	No	Yes	Yes
OR	Unlicensed in the local market	No	Yes	Yes
OR	Confirmed falsified by market authorisation holder	No	No	Yes

The Indonesian regulator BPOM uses the expanded definition, with the addition of any violations related to product information provided on packaging or patient information sheets. We did not have reference packaging, and were thus unable to include this parameter.

Weighted estimates

We weighted the raw prevalence data using product-level sales volume data for calendar 2022 bought from pharmaceutical data aggregator IQVIA, as well as transaction volume data for calendar 2021 provided to us by the national public procurement agency.

IQVIA bases estimated sales volume by medicine, brand, dose and formulation (but not market authorisation number) on data collected from a nationally-representative panel of public and private hospitals, and retail outlets. It provides data for hospitals and retail outlets separately, in smallest counting units (tablet, capsule or 5ml dose). IQVIA data do not capture volumes dispensed in primary health facilities.

While the company provides data by medicine, formulation, and brand or market authorisation holder, authorisation holder data are not available for a substantial proportion of unbranded generics (31% by volume over the five study medicines), often because hospital procurement systems do not record this information. IQVIA reports these volumes under the aggregate heading of "generic manuf".

Primary sector volumes are recorded in the public procurement data, by market authorisation holder. All public hospitals (n=1018 at December 2022), and any private hospital that collaborates with JKN (n=1,945)[21] may also buy medicines for insured patients through the public procurement platform.

We produced estimates weighted by national market size as follows:

Prevalence by product (active ingredient, formulation, brand and manufacturer)

- For products collected in our study, we combined all the products that were of the same medicine, brand, formulation and manufacturer, regardless of dose, and re-calculated the raw prevalence by dividing the total number that tested out of specification on any test by the total tested. (n=284).
- For registered products found in 2022 IQVIA data (and thus clearly available in the market) but not sampled in our study, we imputed prevalence based on available data about most-similar products, using the systematic, iterative process described in the supplementary methods file, Table A15 (n=178)
- For comparative analysis, we repeated this process using the expanded and maximum definitions shown in Table 4.

Volume by product (active ingredient, formulation, brand and dose)

- For products that were not winners of public procurement auctions, we used the IQVIA annual sales volume data (n=374).
- In 2022, 96% of Indonesia's 3072 hospitals were reported to accept publicly insured patients.[21,38] If they used the brand available on the single-winner public procurement platform, they were highly likely to have bought it through that platform. To avoid double counting, we omitted the IQVIA hospital volumes for these products (n=17), on the assumption that these volumes would already be captured in the public procurement data. For these products, the volume was assumed to be the IQVIA retail volume plus the public procurement volume.
- For products found in the study sample (thus clearly available in the market) but not captured in IQVIA data, we assigned the average volume of all products listed in IQVIA but not found in our sample. The underlying assumption was that products captured in one sample but missing in the other are similarly rare/low volume (n=71).
- We reassigned the volume of unbranded generic products for which IQVIA provides no company data (market authorisation holder "generic manuf.") in proportion to market distribution of all unbranded generics for that medicine and dose.
- We adjusted volumes to account for online sales which did not originate from a licensed bricks-and-mortar pharmacy, as described in supplementary methods.
- If an individual product had the same market authorisation holder but two different manufacturers, we split the total volume evenly between the manufacturers.

Adjusted estimates

- For each product (medicine, authorisation holder, brand, dose, formulation and manufacturer), we multiplied the prevalence by the sales volume, to get the total estimated number of out-of-specification units.
- We summed the out-of-specification units and divided them by total sales volume to get an estimate of the actual prevalence of out-of-specification products in circulation nationwide.
- For comparative analysis, we repeated this process using the expanded and maximum definitions shown in Table 4. However, if a regulatory violation was

only found in medicines sampled from unregulated internet sellers, we applied the relevant prevalence only to the estimated volumes for that channel, while continuing to apply the aggregate prevalence found in samples from all other sources to sales through all other channels.

Figure 1 provides a schematic illustration of the imputation process.

To investigate the relationship between product volume and failure rates, we merged product volumes for each product sampled in STARmeds back into our study dataset, and compared frequency of failure by decile of product volume.

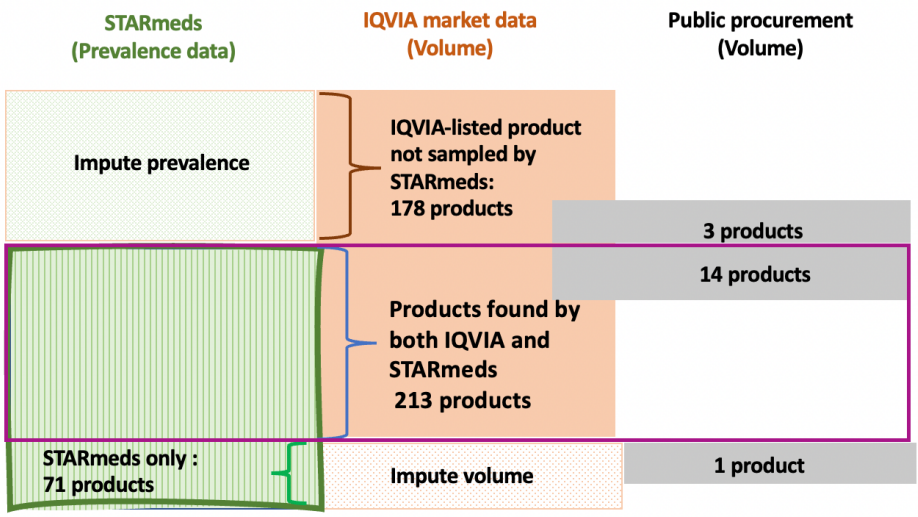

Figure 1: Schematic illustration of imputation process

Ethics and reporting

The study design and purpose were widely discussed with the national regulator and a multisectoral national working group on medicine quality during the design phase. The study protocol was approved by institutional review boards at Universitas Indonesia (970/UN2.F1/ETIK/PPM.00.02/2020, extended by S-736/UN2.F1/ETIK/PPM.00.02/2021) and Imperial College London (21IC7265). We also explained the aims and methods of the study to the district governments of all districts in which samples were collected, and obtained a letter of approval from each.

If we suspected a sample of falsification, we notified BPOM. If a suspect sample came from a hospital or health care provider, we suggested they protect patients by quarantining the product, awaiting confirmation of quality. All sample-level results were shared with BPOM once certificates of analysis were issued; after the deadline for verification of batch numbers and expiry dates had passed, we shared test results of their own products back to responding market authorisation holders.

Protocols were in place in case safety or ethical issues arose in the field. Individual workloads were increased because rigorous COVID-19 screening reduced the number of field staff able to attend training or collect samples; active mystery shoppers were supported by dedicated study staff in the district research hubs, who provided full-time problem-solving support by phone. No other safety or ethical issues arose in the field, beyond the usual logistical challenges (mostly related to seasonal flooding).

Results

We logged a total of 1333 samples, and tested 1274 of them. The remainder of the paper discusses the samples that underwent laboratory testing, unless otherwise specified. Of the tested samples, 82 were of doses different from those targeted in the study.

On the internet, we bought two samples of products which do exist in the Indonesian market but which were in these cases packed for other markets (one, a version of Pfizer's Norvasc/k brand of amlodipine, was labelled in Turkish; the other, GSK's Zyloric brand of allopurinol was of unknown origin). All the other medicines sampled were made in Indonesia, by a total of 72 different manufacturers. Four locally-made branded products were registered by multinational companies; in addition to the two mentioned above, Merck sells amlodipine under the brand name Amcor, while Teva's generics division Allergen/Actavis sells cefixime, marketed as Abixim.) The rest were registered by 75 different Indonesian pharmaceutical firms.

Only one brand, Norvask, was an originator product. The other 179 unique branded products were technically generic medicines (marketed after the patent on the originator product for the active ingredient had expired); we refer to all these products as branded generics. The remaining 101 unique products were identified by their active ingredient, using the International Non-proprietary Name. In accordance with Indonesian regulations, all also printed the name of their market authorisation holder and (if different) their manufacturer on the primary packaging. A total of 22% of unique products were made by companies that differed from their market authorisation holder, indicating contract manufacture.

Table 4 shows the distribution of tested samples by sampling area and medicine; it also provides the number of unique products (different brands/ registered unbranded products for each medicine, dose and formulation). Table 5, meanwhile, shows the distribution of samples by location and type of outlet, giving both the total number of tested samples collected at each outlet type, and the number of unique products.

Overall, we sampled 280 different products (by medicine, dose, formulation and brand) four were made by more than one manufacturer, for a total of 284 unique products. Of these, 182 (64.1%) were branded.

Of the products and doses targeted by our study, we collected a median of 2 and mean of 5.2 samples per product, with significantly higher numbers of samples per unbranded generic, compared with branded medicines (mean 6.7 vs 4.4, $p = 0.008$).

Branded medicines dominated the online samples (72%), reflecting our search strategy, while in physical outlets just over half of samples were branded (51%).

Table 4: Number of samples tested by location of collection, medicines and dose, and number of unique products collected by medicine, dose and branded status

	Number of samples									Number of different products for each active ingredient			
	Greater Jakarta	North Sumatra		East Java		NTT		Online	**Total**	API Total	Branded	Unbranded	Unique products
Medicine & dose		Rural	Urban	Rural	Urban	Rural	Urban						
Allopurinol 100mg	43	14	26	20	31	12	19	55	220		19	16	35
Allopurinol 300mg	21	0	12	3	11	2	4	24	77	**297**	17	5	22
Amlodipine_10mg	3	0	4	0	4	1	5	15	32		8	13	21
Amlodipine 5mg	55	13	25	18	27	7	21	46	212	**244**	28	27	55
Amoxicillin 500mg	44	11	29	22	28	11	22	57	224		32	7	39
Amoxicillin, dry syrup	22	5	17	2	9	2	8	10	75	**299**	15	9	24
Cefixime_100mg	38	9	20	15	20	10	15	41	168		21	7	28
Cefixime_200mg	7	2	3	1	0	1	2	11	27	**195**	6	7	13
Dexamethasone_0.5mg	42	12	27	19	34	10	18	58	220		25	10	35
Dexamethasone_0.75mg	4	1	2	2	0	0	0	10	19	**239**	8	0	8
Total	**279**	**67**	**165**	**102**	**164**	**56**	**114**	**327**	**1,274**		179*	101**	**280**

*Excludes two illegally imported versions of locally registered products

**Unbranded generics from the same market authorisation holder are counted separately for each active ingredient and dose/formulation

Table 5: Distribution of samples and unique products by outlet type and location

Physical outlets													
		Pharmacy		OTC medicine shop		Primary health centre		Hospital		Doctor		Midwife	
		Samples	Products	Samples	Products	Samples	Products	Samples	Products	Samples	Products	Samples	Products
Greater Jakarta		194	90	48	44	6	6	21	18	8	7	2	2
North Sumatra	Medan	125	79	1	1	5	5	10	10	14	11	10	10
	Labuhan Batu	25	23	3	3	6	6	12	12	10	10	11	11
East Java	Surabaya	118	55			2	2	21	15	13	12	10	10
	Malang district	69	51			5	5	14	14	5	5	9	9
NTT	Kupang city	65	40			9	7	21	16	19	17	0	0
	TTS	36	22			4	4	14	11	0	0	2	2
Total		632	165	52	48	37	25	113	54	69	49	44	34

Online									
Regulated				Semi-regulated		Unregulated		**Total**	
Geo-positioned app		Registered online medicine sales site		On-line sales from verified pharmacy					
Samples	Products	Samples	Products	Samples	Products	Samples	Products	Samples	Products
80	62	17	14	44	37	186	123	327	177

OTC: Over the counter
NTT: Nusa Tenggara Timur; TTS: Timor Tengah Selatan

Raw quality estimates

The results of pharmacopeial testing by active ingredient and formulation are shown in Table 6. This gives the number tested and the number failing for each test type, the numbers failing to comply with additional regulatory specifications (displaying a market authorisation number and expiry date valid at the time of purchase) and the overall prevalence of failure. We also report the number confirmed falsified by market authorisation holders, though we note that confirmation was only provided for 542/1274 tested samples. Sample level data, with granular pharmacopeial test results, can be downloaded from the study archive for more detailed analysis.[19]

Overall, 8.2% of samples failed at least one pharmacopeial test. If we used the expanded specifications closest to those used by the Indonesian regulator, an additional 6 samples qualified as out of specification and the total prevalence reached 8.7%. Using our "maximum" definition of poor quality added 13 specimens confirmed falsified by companies which had not failed any other test, taking unadjusted prevalence to 9.7%.

Prevalence differed by medicine type. The anti-hypertensive medicine amlodipine, the only chronic disease medicine in the study, had the lowest testing failure rate, at 1.6% (all failing in uniformity of content). A further two samples of amlodipine brands circulating in Indonesia, both purchased from unregulated internet sellers, were illegally imported from other countries. The antibiotics amoxicillin and cefixime had the highest testing failure rates (10.0% and 19.0% respectively, totalling 13.6%, compared with 4.9% for non antibiotics, $p<0.000$). One in five amoxicillin dry syrup samples failed assay testing; there was no significant difference between tablet and capsule formulations.

Of 72 manufacturers, 30 (41.7%) made at least one of the samples that failed a pharmacopeial test in our study, while at least one out-of-specification product was registered to 37 of the 79 market authorisation holders (46.8%).

A total of 47 samples (3.7%) contained at least some medicines that had passed their expiry date at the time of the last test performed. Samples expired at last testing date were no more likely to fail testing than unexpired samples (8.5% vs 8.2%, $p=0.95$).

Most of the failures were clustered relatively close to permissible limits denoting quality. Two thirds of the failed specimens tested within five percentage points of the permissible limits for assay and dissolution, or five points of the acceptability limit for uniformity of content. 15 samples (1.2% of all tested samples), deviated from permissible limits by 10 percentage points or more (or acceptability points, for uniformity); only 1 of these was confirmed falsified by its market authorisation holder. Figure 2 shows the deviation above (for assay and uniformity) or below (for assay and dissolution) the acceptability limits for each molecule (the "in specification" zone varies between medicines as described in Table 3; it is depicted schematically here as the zone between the dotted lines.)

Table 6: Pharmacopeial test results by medicine type

Medicine & formulation	Assay			Dissolution			Uniformity of content			Other regulatory violation					Totals						
															N	Core		Expanded		Maximum	
	N	OOS	% Fail	N	OOS	% Fail	N	OOS	% Fail	No ED	Exp.	No NIE	Falsified N	Falsified CF		OOS	% Fail	OOS	% Fail	OOS	% Fail
Allopurinol tablets	297	13	4.4	283	8	2.8	0	-	-	1	1	1	107	5	297	18	6.1	21	7.1	26	8.8
Amlodipine tablets	244	0	0.0	236	0	0.0	83	4	4.8	1	0	1	87	2	244	4	1.6	6	2.5	7	2.9
Amoxicillin capsules	36	0	0.0	32	0	0.0	0	-	-	0	0	0	7	0	36	0	0.0	0	0.0	0	0.0
Amoxicillin dry syrup	75	15	20.0	0	-	-	0	-	-	0	0	0	25	0	75	15	20.0	15	20.0	15	20.0
Amoxicillin, tablets	188	3	1.6	187	12	6.4	0	-	-	0	1	0	87	4	188	15	8.0	16	8.5	20	10.6
Cefixime capsules	178	11	6.2	165	36	21.8	0	-	-	0	0	0	100	7	180	32	17.8	32	17.8	34	18.9
Cefixime tablets	17	4	23.5	9	1	11.1	0	-	-	0	0	0	6	2	15	5	33.3	5	33.3	5	33.3
Dexamethasone tablets	239	11	4.6	228	9	3.9	93	2	2.2	0	0	0	123	1	239	16	6.7	16	6.7	17	7.1
All	**1274**	**57**	**4.5**	**1140**	**66**	**5.8**	**176**	**6**	**3.4**	**2**	**2**	**2**	**542**	**21**	**1274**	**105**	**8.2**	**111**	**8.7**	**124**	**9.7**

NIE: Nomor Izin Edar, Market authorisation number
ED: expiry date
CF: Confirmed falsified by market authorisation holder. A total of xx Market Authorisation holders provided confirmation data. Here we give the total number with any confirmation of status, and the number confirmed falsified

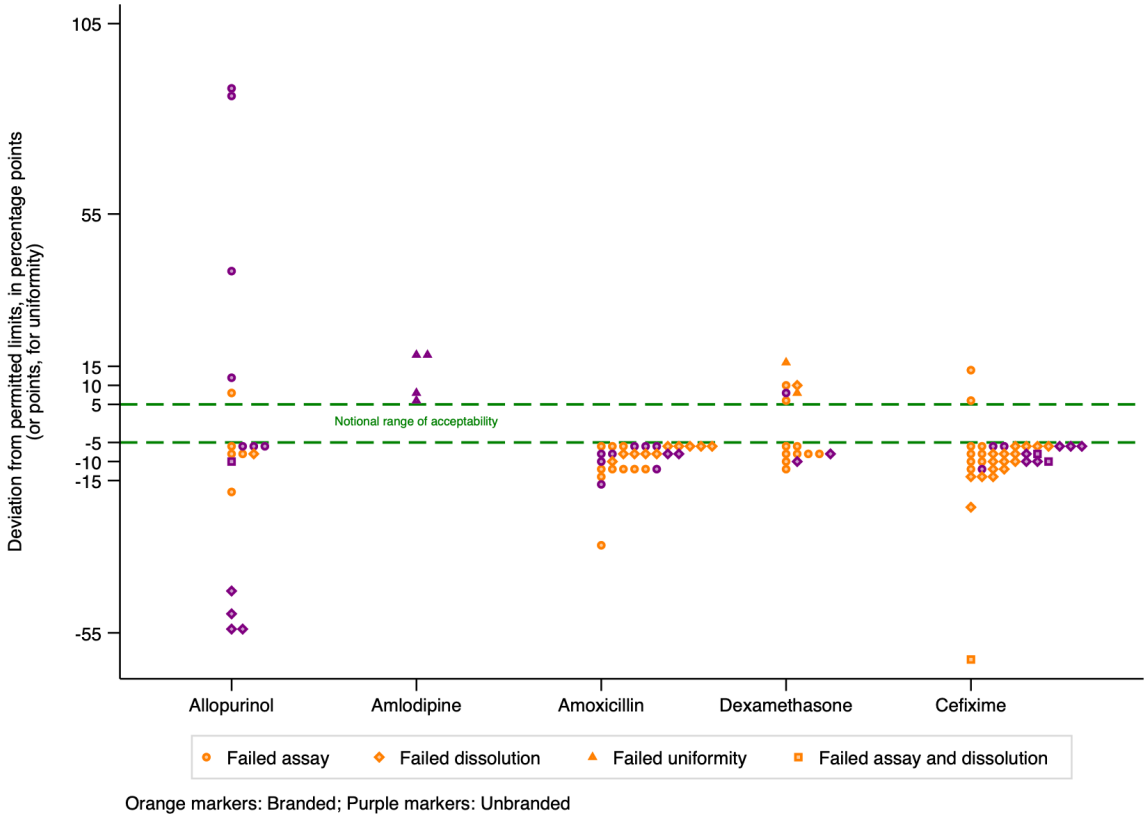

Figure 2: For all samples failing pharmacopeial testing, maximum deviation on any test from permitted limits, by active ingredient

Falsified medicines

All 79 market authorisation holders were contacted on a minimum of 3 occasions to confirm that sample details accorded with their manufacturing records. 44 had provided confirmation by the time of writing (October 1 2023). The confirmations covered 569 samples, including 27 which were not laboratory tested -- none of the latter were falsified.

12 companies reported that data for at least one product was not in accordance with manufacturing records. 21 of the 569 samples with confirmed data were reported falsified (3.7%) A third of falsified samples failed in pharmacopeial tests, compared with 9.4% of those with correct production records ($p < 0.000$)

Of the 21 falsified products, 6 had fake batch numbers, while 15 had incorrect expiry dates. Only one falsified product was an unbranded generic (prevalence of falsified generics by branded status (6.2% vs 0.2%, $p < 0.000$).

Over half of the falsified samples (11/21) were flagged as suspicious at visual inspection by study staff. Of samples confirmed by authorisation holders, prevalence of falsified products among those flagged as suspicious was 20%, compared with just 2% in unflagged samples ($p < 0.000$)

The majority of falsified samples (16/21) were purchased on the internet, all but 1 from unregulated sellers. The remainder were from an over-the-counter medicine shop, a doctor and a midwife, and pharmacies (2), one of which was in an unregulated wholesale market. Only 1 falsified sample was acquired from an outlet type that BPOM

might include in its routine post-market surveillance for prescription medicines. That outlet was in the most remote district in our study.

Source of medicines

In regression analysis, after controlling for differences in medicines, the odds of testing out of specification were 2.2 times higher for samples collected in rural areas, compared with those collected in cities ($p<0.001$). There was no significant difference in the pharmacopeial quality of samples bought from different types of outlets, including those bought online compared with those acquired from physical sources (8.0 vs 8.3%, $p=0.83$). However, among samples acquired online, those bought from individuals selling on general marketplaces or social media were more likely to fail any pharmacopeial test than those bought from licensed online vendors or verified online stores of bricks-and-mortar pharmacies (11.3 vs 3.6%, $p=0.01$). This relationship held after controlling for differences in medicines; in that case, people buying medicines from unregulated online sellers were 3.7 times more likely to get a substandard product compared with those buying from a regulated or semi-regulated online seller ($p=0.013$).

In terms of likelihood of failing any pharmacopeial test, we found no significant difference between samples bought from the types of outlets included by the regulator BPOM in routine post-market surveillance, compared with excluded outlets (8.9 vs 7.8%, $p=0.46$). However, if we used the maximum definition of poor quality, including other regulatory violations and confirmed falsification, products bought from outlet types not sampled by the regulator were significantly more likely to fail (11.8% vs 8.2%, $p=0.03$).

Branded status and price

There was no significant difference in quality between branded and unbranded products (9.1% vs 7.2% on pharmacopeial tests alone, $p=0.23$) including after controlling for differences in medicine, district, or source. However, if we use the expanded definition of poor quality, including unregistered and expired products and those confirmed falsified, then branded products were more likely to be categorised as poor quality than unbranded generics (11.6% vs 7.4%, $p=0.012$). Medicines available free to patients in the public insurance system were significantly less likely to fail testing than medicines paid for out of pocket (4.1% vs 9.0%, $p=0.23$).

Prices for the same medicine, dose and formulation varied widely between brands (and between manufacturers of unbranded versions), and will be reported in detail elsewhere. Figure 3 summarises the variation in prices, comparing branded with unbranded generics. For each medicine, dose and formulation, we calculated median prices across all samples (including those not tested), valuing medicines that were given to patients for free at the public procurement price plus allowed mark-up. We then compared the price of each sample to that benchmark price. On average, people buying branded medicines paid 61% over the benchmark price, with a mean of 3.6 times the benchmark (standard deviation 3.5). Unbranded medicines traded in a much narrower price range.

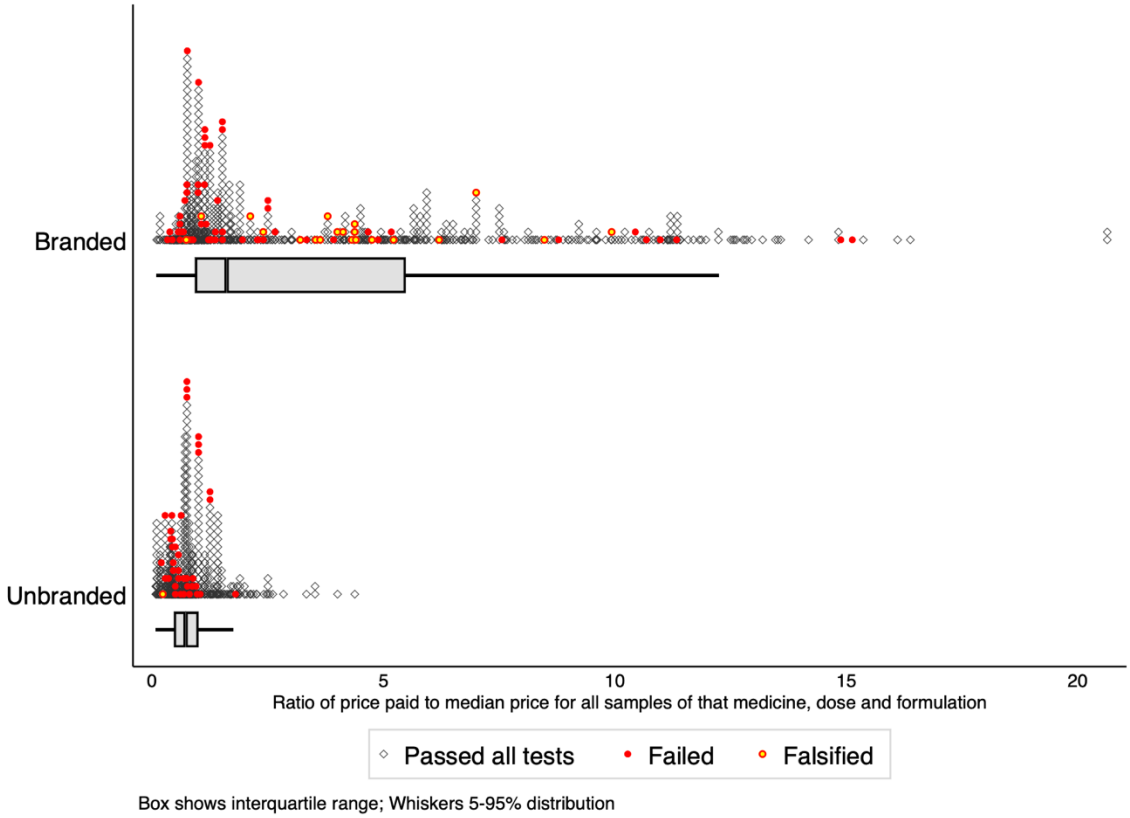

Figure 3: Variation of prices paid for branded and unbranded medicines

In logistic regression, there was no relationship between price and pharmacopeial quality, including after controlling for differences in medicine, district, or source.

Samples confirmed as falsified sold at an average of 4.1 times the median price for the medicine, dose and formulation, but there was no significant difference in price between falsified and non-falsified products of the same brand.

Volume-adjusted quality estimates

Following the imputation and estimation procedures described in the methods, we compared the raw prevalence of medicines failing any pharmacopeial test in our study with the prevalence adjusted for market size.

If we disregard different dosages, STARmeds sampled 232 unique products (by active ingredient, brand, manufacturer and formulation). Figure 4 plots the measured prevalence of these products against the quintiles of market volume for each, indicating also how many samples of each product we tested, independent of dose). The number of samples tested per unique product ranged from 1- 59.

Of 232 unique products, 170 (73.3%), including 90.4% of those tested only once, had no testing failures among the samples collected. However, test results were not consistent across different samples of the same product. Almost three quarters of the 34 products of which we tested more than ten dose-agnostic samples (shown as the green diamonds in Figure 4) fell into the highest quintile for sales volume, and 62% failed no tests. Prevalence of failure in the other frequently-sampled products ranged from 1.7%

to 61.5%. The per-product failure rates for all 232 dose-agnostic products are shown in Supplementary Figure 1.

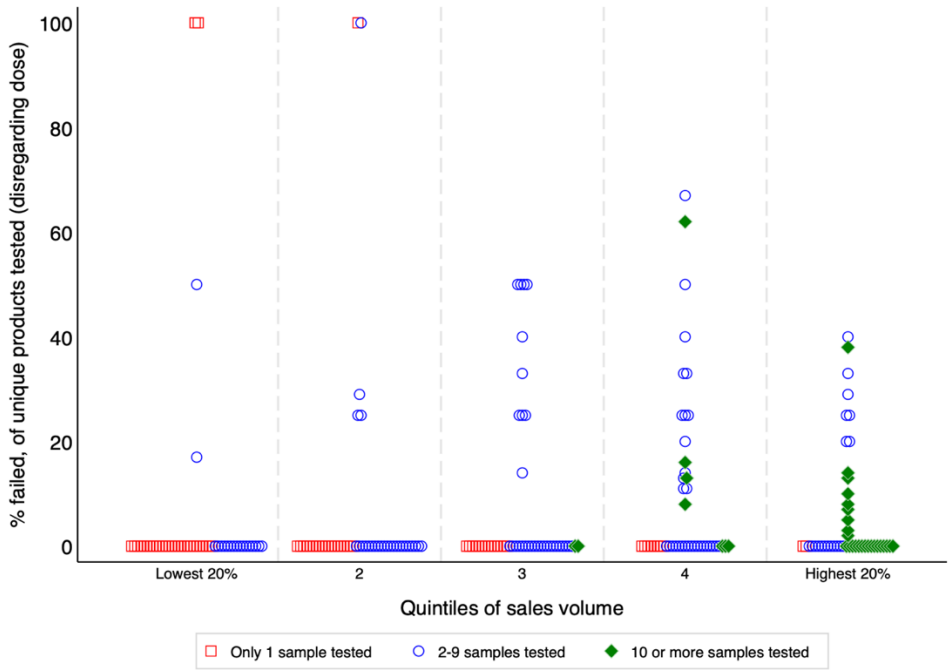

Figure 4: Estimated prevalence of poor quality samples, by market volume

All four products estimated to be entirely of poor quality (with 100% percent failing at least one test) were in the lowest two quartiles for volume; 3/4 were tested only once.

Among our 1274 study samples, the odds of testing out of specification fell significantly as sales volume rose. In logistic regression, after controlling for active ingredient, the odds of testing out of specification decreased by 15% for every increasing decile of volume ($p<0.000$).

Figure 5 compares prevalence of out of specification medicines measured in the samples collected in our study (raw core prevalence) with the estimated prevalence after adjusting for market share of each product. We also show the estimated failure rate when imputed values for brands not sampled in STARmeds fieldwork are included.

Adjusted prevalence of out of specification samples across products tested in the study was 40.3% lower than raw prevalence (4.9% vs 8.2%). Adjusted prevalence was lower for every medicine and formulation except for amoxicillin dry syrup, which rose by 8.2%. This resulted from our imputation strategy, which calculated prevalence across all samples of that active ingredient, formulation and brand, regardless of dose. The strategy was adopted to increase robustness of estimates for non-target doses, of which samples were scarce. In this case, 3/8 samples of the target dose (125 mg/ml) of a specific brand failed, compared with 0/1 of the non-target dose (250 mg/ml). The total prevalence of 33.3% (3/9) was thus applied to the large volumes of 250 mg/ml of this product sold in the market estimates, causing an overall rise for the estimated volume of substandard amoxicillin in this formulation.

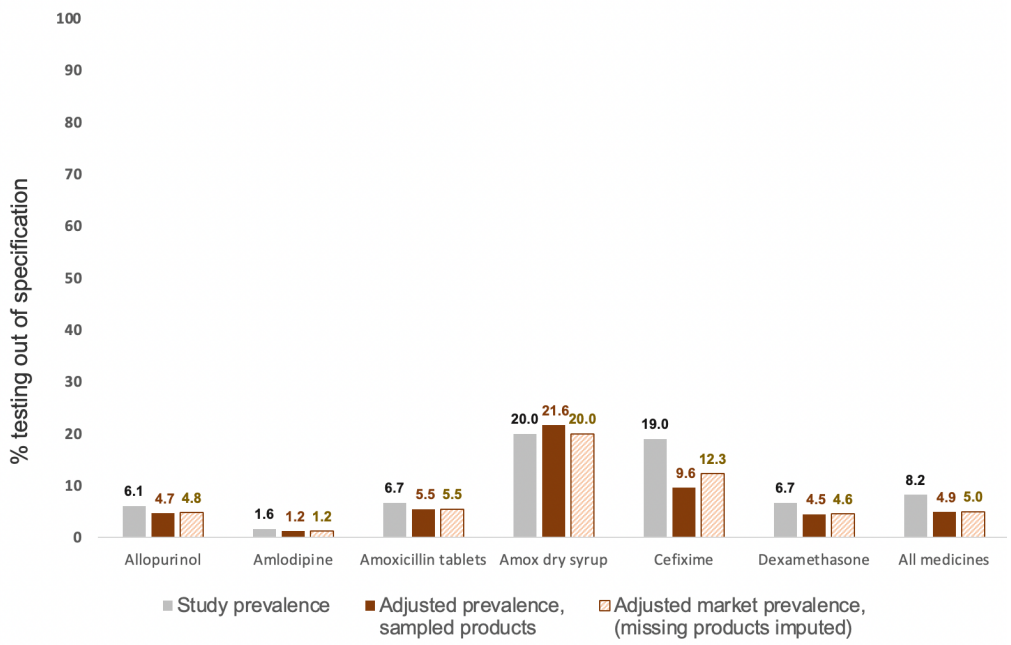

Figure 5: Raw and adjusted prevalence for study medicines, with adjusted market prevalence

For other products, using the same procedures, the relative difference ranged from a drop in prevalence of poor quality products of 18% for amoxicillin tablets to a drop of 49.5% for cefixime. When we added estimates for products not sampled in our study (imputing prevalence based on manufacturer and other characteristics as described in the methods supplement), overall market prevalence was 5.0%, a 39.3% drop from raw prevalence.

In the raw estimates, the expanded and maximum definitions of poor quality increased overall prevalence by 0.5 and 1.5 percentage points when compared with the core definition that considered only pharmacopeial testing. These figures represent increases of 5.7% and 18.1% in the overall estimates prevalence of poor quality products. However, once adjusted for market volume, the differences were much smaller (0.1 percentage points for the expanded definition and 0.4 for the maximum definition for sampled products, equivalent to differences of 1.5% and 7.8% in the estimated prevalence of poor quality products; see supplementary Figure 2). This was in large part because the many of the falsified products were only sampled from unregulated internet sellers, where estimated sales volumes are very small.

When translated back to 3.38 billion units of STARmeds-sampled products sold annually in the Indonesian market, the difference between raw and adjusted prevalence represents a difference of 111 million tablets, capsules or 5ml units of medicine a year.

Discussion

The raw prevalence of out of specification products measured in our survey of 1,274 samples of five medicines from a wide variety of outlets in different areas of Indonesia was 8.2%. However, when we adjusted for market size the estimate fell by 40.3%, to 4.9% (5.0% when imputing prevalence and volumes to account for products not sampled). We found that poor quality medicines do not, in general, achieve large market shares.

In common with other field surveys in lower-income settings,[39,40] we identified significant differences in quality between medicines. No amlodipine sample failed assay or dissolution testing, a result consistent with findings in an earlier study of cardiovascular and diabetic medicines conducted in just one district of Indonesia.[18] At the other extreme, one in five of the 75 samples of amoxicillin dry syrup contained below the permitted amount of active ingredient. Failure on both assay and dissolution were also high for cefixime, at 7.7% and 14.4% respectively.

Most failed samples hovered close to permitted limits (two thirds were within five percentage points of allowed limits for assay or dissolution). The impact for individual patients may thus be limited. However, the danger posed by minimally substandard antibiotics (with assays or dissolutions close to the lower limits of acceptability) may be more considerable. Slightly lower than permitted dissolution, particularly, may place the concentration of active ingredient within the mutation selection window, with enough potency to kill susceptible strains of a bacteria but not mutations, favouring the development and spread of bacteria resistant to antibiotics.[5,41] This is especially worrying in the case of cefixime, which is assigned by WHO to the "Watch" category of antibiotics, prone to resistance and prioritised for stewardship programmes.[42]

Our adjusted estimate for five medicines was 23% higher than the 4.0% found by the Indonesian regulator BPOM in post-market surveillance across all medicines in the months before the study.[43] STARmeds sampling strategy and definitions of quality did not mirror BPOM's exactly. We randomised at the level of the outlet in just seven districts and the capital region, testing fewer than 1,300 samples collected in early 2022, while BPOM's largely random sampling is designed to be representative of products at the national level, and included of 13,539 samples collected in calendar 2021.

The differences in sampling could affect the results in both directions. Within our sampling districts, STARmeds sampling was more comprehensive in terms of outlets included. In addition to the pharmacies, hospitals and clinics from which BPOM collects prescription medicines, we sampled from the internet, over-the-counter medicine shops, wholesale markets, doctors and midwives -- virtually all the sources from which Indonesian patients get prescription medicines. We also used mystery shoppers in retail outlets, while the regulator samples overtly. We would expect both distinctions to *increase* the likelihood that we would collect poor quality products, compared with BPOM's strategy.

Other differences are more likely to *decrease* our chances of sampling poor quality products, compared with BPOM. The STARmeds study did not sample sterile formulations, or those with a narrow therapeutic index, which have more stringent specifications and can thus more easily fail testing. For reasons of feasibility, we also excluded locally rare and psychoactive medicines, and expensive medicines not covered by insurance, all of which may be at greater risk for falsification. In addition, BPOM includes in its definition of out-of-specification any product whose packaging or product information leaflets differ from those submitted at the time of registration. We did not have reference packaging for comparison, so could not include samples that failed only on this measure in our estimates.

In our study, we compared three different definitions of poor quality: a measure based simply on pharmacopeial testing; a measure including all other regulatory parameters considered by BPOM except packaging, and a measured which included samples confirmed falsified. The differences in estimated prevalence using the three measures

was significant in our raw data, in part because we invested considerable effort in sourcing samples from unregulated online sellers, who dominated the sales of samples confirmed as falsified. However, when adjusted for market volume, the differences were greatly reduced. To increase comparability between field surveys of medicine quality, we thus recommend reporting results using only pharmacopeial definitions of quality, which best reflect the immediate threat to patients.

Studies such as this one summarise quality in prevalence figures based on pass/fail criteria. A binary distinction that classifies a medicine with 89.99% of the labelled amount of active ingredient as poor quality and one of 90.01% as good quality -- while practical for regulatory purposes -- may be of limited use in understanding the extent of the threat to patients, or in planning a proportionate response. This is especially true since there are now over 60 different pharmacopeia in the world.[44] They do not all set acceptability thresholds at the same level and none, to our knowledge, publishes the data on which their choice of threshold is based. Some researchers have tried to distinguish between minor and extreme deviations when reporting results, though again without clarifying the basis for their chosen limits.[40,45]

Given the differences in definitions and sampling, the weighted estimate in our study is quite similar to that published by the national regulator for the Indonesian market as a whole, yet it was achieved with a sample that was less than one tenth of the size. A full consideration of the cost of surveillance will be published elsewhere, but we suggest that weighting raw estimates by market size is likely to be a cost-effective option for arriving at a better understanding of the actual magnitude of the threat posed by substandard or falsified medicines in national markets. Sources of data on sales volumes vary by country, and data from commercial aggregators, where available, are expensive. Other potential sources include volumes recorded by distributors, wholesalers, public procurers and public insurers, as well as tax and customs authorities.

Our findings are consistent with the one other study we are aware of which has adjusted quality estimates for market size.[17] In that study of an antimalarial, an antibiotic and an analgesic in the Democratic Republic of Congo, the raw estimates of poor quality medicines was 27.2%, falling to 1.3% when adjusted for market size data provided by medicine distributors.

Our findings were also comparable with those in other countries in terms of price. In common with one other study of medicine quality in Indonesia and nations as diverse as Cambodia, India and Togo, we found that quality did not vary by price.[40,46,47] Since the introduction of Indonesia's national health insurance scheme, which is expected to provide free medicines to any member willing to follow the correct procedures (which often involve long waits in crowded public facilities), Indonesian media have carried reports questioning the quality of free medicines.[48,49] Research in Indonesia and other countries (including the United States, China and South Africa) have similarly identified distrust of the quality of free medicines and unbranded generics, among both patients and health-care providers. Researchers sometimes suggest that distrust is deliberately sown to favour the financial interests of doctors, hospitals and pharmaceutical companies.[50–54,27] Our study found no basis for distrust of free medicines or unbranded generics.

More surprisingly to us, there was a limited association between where medicines were sold and their quality. Out of specification medicines were collected from every outlet type, including both public and private hospitals, and in every district. Failure rates

were, however, higher in rural areas, perhaps reflecting degradation in the supply chain. Medicines sold online were similar in quality to those collected from bricks-and-mortar sources. However, if shoppers went to the often considerable effort of buying from online sellers not associated with verified bricks-and-mortar pharmacies, the products delivered were significantly more likely to be out of specification or falsified.

Only 1 confirmed falsified product was found in a regulated outlet permitted to dispense prescription medicines. We invested considerable time in engaging with manufacturers to verify production records, eventually getting response from 44 of 79 companies. Unexpectedly, the majority of the samples that were confirmed falsified by manufacturers passed quality testing; the majority also had expiry dates which did not match genuine batch numbers. This suggests that many products targeted by falsifiers continue to maintain serviceable quality beyond the expiry dates permitted in the authorisation paperwork. We note that simple visual inspection by researchers familiar with brand packaging identified more than half of the products that were confirmed falsified. A pilot programme to train health care workers in Indonesia and Tanzania to spot and report suspect medicines was deemed promising and could be expanded.[55] However the signs would not be evident to most consumers, who do not have the same points of reference.

We recognise that there are other ways that these data could be examined, for example looking in more detail at differences by geographic location, source, time to expiry and other factors. Our team is planning more detailed investigation of quality by price variation and by different types of online outlets, as well as enquiries into compliance with transparency and pricing regulations, and the costs of post-market surveillance. We have deposited in our study repository the individual sample level data (with company names and some other identifiers masked in accordance with the terms of our ethics approval), as well as an English language codebook. We encourage their re-use by other researchers with an interest in the quality of medicines in a large and diverse middle-income market.

We sampled medicines across four therapeutic groups from all sources from which Indonesian patients commonly acquire medicines, and performed four major pharmacopeial tests (identification, quantification and uniformity of content, and dissolution). An important limitation of our study, along with most other academic studies and much post-market surveillance, was that we could not afford to test for impurities. Even relatively well resourced regulators, including BPOM, do not commonly test for impurities, or for by-products of non-active ingredients. This explains why the Indonesian regulator did not pick up the presence of lethal non-active ingredients in the paediatric syrups during their extensive routine post-market surveillance. It underlines the need for more affordable medicine testing technologies, as well as greater attention to impurities and non-active ingredients in considering the quality of medicines.[56–58]

Its limitations notwithstanding, our study remains one of the most comprehensive single-country academic surveys of medicine quality. And yet we underline that the field survey data imperfectly reflect the prevalence of out of specification products in the market, or their true threat to public health. When we adjusted for brand volumes, including medicines provided for free in the public procurement system, we found considerably lower prevalence of out-of-specification products than the "headline" figure from our study suggests.

In common with authors of recent systematic reviews, we believe the aggregation of results across studies which do not publish sample-level data, which use varying definitions of quality, and which do not consider market distribution of out-of-specification samples to be misleading.[3,5–7] We hope our work encourages regulators and others conducting post market surveillance or primary surveys of medicine quality to consider weighting by market size as a cost-effective tool for improving understanding of the volume and distribution of poor quality products, as well as to advocate for adequate investment in quality assurance in production and oversight of the supply chain.

We note, however, that even estimates weighted by market volume are expensive and time-consuming to generate. Health authorities must assess the value of this investment relative to a case-finding approach with narrower focus on identifying and removing substandard and falsified products at highest risk of harming public health.[10] One option would be to conduct market-wide surveys of prevalence every four or five years, with more targeted, risk-based surveillance in interim years. We suggest that these strategies in combination could deliver the most cost-effective balance of data necessary to plan and calibrate regulator responses and minimise risk to patients of substandard and falsified medicines.

Data availability statement

Additional data are available in three locations, all within the STARmeds repository. Supplementary data for **this specific paper** (including the product volume data file used for adjusted estimates, the analysis code in Stata format for this paper, the supplementary methods description, supplementary figures and author contributions) are at: https://doi.org/10.7910/DVN/QRKDWG

Data and documentation related to STARmeds fieldwork more generally are in the study archive. This archive is easiest to use in Tree view. It contains **the sample level data** produced by the STARmeds field study, including raw laboratory data, in csv format. Also included are laboratory protocols and a more detailed description of methods. The archive can be accessed at: https://doi.org/10.7910/DVN/RKYICP.

Finally, we provide a **free Toolkit** to help researchers and regulators design and implement medicine quality field surveys using mystery shoppers. The toolkit contains downloadable and adaptable versions of data collection software, field control forms, field worker contracts and other potentially useful documentation. The Toolkit can be downloaded from: https://doi.org/10.7910/DVN/OBIDHJ

Conflict of interest statement

Yusi Anggriani is a member of the Indonesian Ministry of Health's advisory committee on medicine pricing, and a member of the World Health Organization Technical Advisory Group on Pricing Policies for Medicines. Elizabeth Pisani has worked as a consultant on research commissioned by the WHO Incidents and Substandard/Falsified medical products team. All other authors report no conflict of interest.

Study Group members, and acknowledgements

STARmeds was a collaboration between Universitas Pancasila, Imperial College London and Erasmus University Rotterdam.

STARmeds study group members listed alphabetically by institution. Group member roles for this paper are provided in the paper-level supplementary materials.

Universitas Pancasila

Yusi Anggriani, Esti Mulatsari, William Nathanial, Yunita Nugrahani, Elizabeth Pisani, Jenny Pontoan, Ayu Rahmawati, Mawaddati Rahmi, Stanley Saputra, Hesty Utami.

Imperial College London

Adrian Gheorghe, Katharina Hauck, Sarah Njenga, Sara Valente de Almeida.

Erasmus University Rotterdam

Amalia Hasnida.

The team would like to thank Indonesia's medicine regulator (Badan Pengawas Obat dan Makanan, BPOM) and the national statistics bureau (Badan Pusat Statistik, BPS) for active support in developing the methods described here. The majority of the data collectors were partners of BPS. We thank them for their hard work. We thank BPS statistical consultants Ardi Adji and Budi Santoso.

We also thank members of a multisectoral working group on medicine quality estimation known as PEMO, which groups 12 Indonesian government institutions and 5 professional or industry associations, for advice provided over the course of the study, as well as the members of our Study Advisory Group (Michael Deats, Kharisma Nugroho, Yodi Mahendradhata, Raffaella Ravinetto, Selma Siahaan, Val Snewin, Virginia Wiseman and Firman Witoelar) for helpful advice.

Funding source

The study was funded by UK taxpayers through the UK Department of Health and Social Care and the National Institute for Health Research, under NIHR Global Health Policy and Systems Research Commissioned Awards, grant number NIHR131145. Thank you.

Thanks also to United States Pharmacopeia, who provided reference standards at discounted prices.

References

1. World Health Organization. A study on the public health and socioeconomic impact of substandard and falsified medical products [Internet]. Geneva: WHO; 2017 Nov [cited 2017 Nov 29]. Report No.: WHO/EMP/RHT/2017.02. Available from: http://who.int/medicines/regulation/ssffc/publications/Layout-SEstudy-WEB.pdf

2. Buckley GJB, Gostin LO. Countering the problem of falsified and substandard drugs: committee on understanding the global public health implications of substandard, falsified, and counterfeit medical products. Washington, D.C: National Academies Press : Institute of Medicine of the National Academies; 2013. 351 p.

3. Do NT, Bellingham K, Newton PN, Caillet C. The quality of medical products for cardiovascular diseases: a gap in global cardiac care. BMJ Global Health. 2021 Sep 1;6(9):e006523.

4. Ozawa S, Evans DR, Bessias S, Haynie DG, Yemeke TT, Laing SK, et al. Prevalence and estimated economic burden of substandard and falsified medicines in low- and middle-income countries: a systematic review and meta-analysis. JAMA Network Open. 2018 Aug 10;1(4):e181662.

5. Zabala GA, Bellingham K, Vidhamaly V, Boupha P, Boutsamay K, Newton PN, et al. Substandard and falsified antibiotics: neglected drivers of antimicrobial resistance? BMJ Global Health. 2022 Aug 1;7(8):e008587.

6. Saraswati K, Sichanh C, Newton PN, Caillet C. Quality of medical products for diabetes management: a systematic review. BMJ global health. 2019;4(5):e001636–e001636.

7. Do NT, Boupha P, Newton PN, Caillet C. The quality of antiretroviral medicines: an uncertain problem. BMJ Glob Health. 2023 Mar;8(3):e011423.

8. McManus D, Naughton BD. A systematic review of substandard, falsified, unlicensed and unregistered medicine sampling studies: a focus on context, prevalence, and quality. BMJ Global Health. 2020 Aug 1;5(8):e002393.

9. US Food and Drug Administration. Postmarket Surveillance Under Section 522 of the Federal Food, Drug, and Cosmetic Act - Guidance for Industry and Food and Drug Administration Staff [Internet]. USFDA; 2022 [cited 2023 Sep 14]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/postmarket-surveillance-under-section-522-federal-food-drug-and-cosmetic-act

10. Pisani E, Hasnida A, Rahmi M, Kok MO, Harsono S, Anggriani Y. Substandard and Falsified Medicines: Proposed Methods for Case Finding and Sentinel Surveillance. JMIR Public Health and Surveillance. 2021 Aug 16;7(8):e29309.

11. Aroca Á, Guzmán J. [Model for a risk-focused approach to health inspection, surveillance, and control in Colombia]. Rev Panam Salud Publica. 2017 Aug 21;41:e105.

12. Fleming Fund. Substandard and Falsified Medicines (WHO) [Internet]. Fleming Fund. Undated [cited 2023 Sep 14]. Available from: https://www.flemingfund.org/grants/substandard-and-falsified-medicines-who/

13. Newton PN, Bond KC, Oxford Statement signatories. Global access to quality-assured medical products: the Oxford Statement and call to action. Lancet Glob Health. 2019;7(12):e1609–11.

14. Slamet L. Policy brief: pengawasan obat dan makanan, termasuk keamanan pangan. Kementrian PPN/Bappenas; 2019.

15. WHO Expert Committee on Specifications for Pharmaceutical Preparations. Guidelines on the Conduct of Surveys of the Quality of Medicines. WHO Technical Report Series, No. 996, 2016, Annex 7 [Internet]. World Health Organization; 2016 [cited 2017 Dec 7]. Available from: https://digicollections.net/medicinedocs/#d/s22404en

16. Newton PN, Lee SJ, Goodman C, Fernández FM, Yeung S, Phanouvong S, et al. Guidelines for Field Surveys of the Quality of Medicines: A Proposal. PLoS Medicine. 2009;6(3):e52.

17. Schiavetti B, Wynendaele E, Spiegeleer BD, Mbinze GJ, Kalenda N, Marini R, et al. The Quality of Medicines Used in Children and Supplied by Private Pharmaceutical Wholesalers in Kinshasa, Democratic Republic of Congo: A Prospective Survey. The American Journal of Tropical Medicine and Hygiene. 2018 Jan 8;98(3):894–903.

18. Dewi A, Patel A, Palagyi A, Praveen D, Ihsan BRP, Hariadini AL, et al. A study of the quality of cardiovascular and diabetes medicines in Malang District, Indonesia, using exposure-based sampling. BMJ Global Health. 2022 Nov 1;7(11):e009762.

19. STARmeds Study Team. STARmeds study repository: Methods, protocols tools and code [Internet]. Harvard Dataverse; 2023 [cited 2023 Aug 8]. Available from: https://doi.org/10.7910/DVN/OBIDHJ

20. STARmeds Field Team. Conducting randomised field surveys of medicine quality using mystery shoppers: A practical guide and toolkit. [Internet]. Harvard Dataverse; 2023 [cited 2023 Aug 8]. Available from: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/OBIDHJ

21. BPJS Kesehatan. Laporan pengelolaan program 2022 dan laporan keuangan 2022 (auditan) [Internet]. Jakarta, Indonesia: BPJS Kesehatan; 2023 [cited 2020 Jul 21]. Available from: https://bpjs-kesehatan.go.id/#/informasi-publik-detail?slug=laporan-pengelolaan-program-dan-keuangan-bpjs-kesehatan-tahun-2022-auditan

22. Anggriani Y, Ramadaniati HU, Sarnianto P, Pontoan J, Suryawati S. The impact of pharmaceutical policies on medicine procurement pricing in Indonesia under the implementation of Indonesia's social health insurance system. Value Health Reg Issues. 2020;21:1–8.

23. Statista. Pharmaceuticals - Indonesia [Internet]. Statista. 2023 [cited 2023 Aug 31]. Available from: https://www.statista.com/outlook/hmo/pharmaceuticals/indonesia

24. Tirtokoesnadi. Industri farmasi perlu perhatian BPJSK, Kemenkes, BPOM, Kemenkeu untuk menghadapi saat Covid-19/virus Corona mewabah dan perkembangan industri farmasi kedepan. GP Farmasi; 2020.

25. Ramadhan A. GP Farmasi usulkan alokasi pembayaran obat dalam anggaran BPJS Kesehatan. Antara News [Internet]. 2019 Jan 19 [cited 2019 Jan 21]; Available from: https://www.antaranews.com/berita/789358/gp-farmasi-usulkan-alokasi-pembayaran-obat-dalam-anggaran-bpjs-kesehatan

26. GP Farmasi. Hutang jatuh tempo obat & alkes JKN belum dibayar mencapai Rp. 3,5 T per Juli 2018. 2018.

27. Hasnida A, Kok MO, Pisani E. Challenges in maintaining medicine quality while aiming for universal health coverage: a qualitative analysis from Indonesia. BMJ Global Health. 2021 May 1;6(Suppl 3):e003663.

28. Ministry of Health Indonesia. Kasus Baru Gangguan Ginjal Akut Pada Anak, Pemerintah Siapkan Langkah Antisipatif [Internet]. 2023 [cited 2023 Aug 30]. Available from: https://sehatnegeriku.kemkes.go.id/baca/umum/20230206/5642343/kasus-baru-gangguan-ginjal-akut-pada-anak-pemerintah-siapkan-langkah-antisipatif/

29. WHO. Full list of WHO Medical Product Alerts [Internet]. Regulation and prequalification. 2023 [cited 2023 Jul 31]. Available from: https://www.who.int/teams/regulation-prequalification/incidents-and-SF/full-list-of-who-medical-product-alerts

30. Badan Pengawas Obat dan Makanan. Informasi Sirop Obat [Internet]. BadanPOM. 2023 [cited 2023 Aug 31]. Available from: https://www.pom.go.id

31. Arlinta D. Perbaikan Sistem Pengawasan Jaminan Mutu Obat Mendesak. Kompas [Internet]. 2022 Oct 31 [cited 2023 Aug 31]; Available from: https://www.kompas.id/baca/humaniora/2022/10/31/perbaikan-sistem-pengawasan-jaminan-mutu-obat-mendesak

32. Ratcliffe R. Indonesian families sue government over deaths from syrup medicines. The Guardian [Internet]. 2022 Dec 2 [cited 2022 Dec 22]; Available from: https://www.theguardian.com/world/2022/dec/02/indonesian-families-sue-government-over-deaths-from-syrup-medicines

33. Parliament of Indonesia SD. Legislator Nilai Peran BPOM Dalam Pengawasan Obat Bersifat Pasif [Internet]. 2022 [cited 2023 Sep 14]. Available from: http://www.dpr.go.id/berita/detail/id/41503

34. Kobo Inc. KoBoToolbox | Data Collection Tools for Challenging Environments [Internet]. KoBoToolbox. 2022 [cited 2022 Jun 29]. Available from: https://www.kobotoolbox.org/

35. United States Pharmacopeia. United States Pharmacopeia National Formulary (43rd edition NF38). USP; 2021.

36. Kementerian Kesehatan Republik Indonesia. Farmakope Indonesia Edisi VI [Internet]. 2020 [cited 2023 Aug 13]. Available from: https://farmalkes.kemkes.go.id/2020/11/farmakope-indonesia-edisi-vi/

37. Stata Corp. Stata. Texas, USA: Stata Corp; 2021.

38. Badan Pusat Statistik. Statistik Indonesia 2023 [Internet]. Jakarta: BPS; 2023 Feb [cited 2023 Sep 15]. Available from: https://www.bps.go.id/publication/2023/02/28/18018f9896f09f03580a614b/statistik-indonesia-2023.html

39. Schäfermann S, Hauk C, Wemakor E, Neci R, Mutombo G, Ngah Ndze E, et al. Substandard and Falsified Antibiotics and Medicines against Noncommunicable Diseases in Western Cameroon and Northeastern Democratic Republic of Congo. The American Journal of Tropical Medicine and Hygiene [Internet]. 2020 May 11 [cited 2020 May 17]; Available from: http://www.ajtmh.org/content/journals/10.4269/ajtmh.20-0184

40. Schäfermann S, Wemakor E, Hauk C, Heide L. Quality of medicines in southern Togo: investigation of antibiotics and of medicines for non-communicable diseases from pharmacies and informal vendors. PloS one. 2018;13(11):e0207911.

41. Pisani E. Antimicrobial resistance: What does medicine quality have to do with it? [Internet]. London: Antimicrobial Review; 2015 Nov [cited 2016 Feb 11]. Available from: http://amr-review.org/sites/default/files/ElizabethPisaniMedicinesQualitypaper.pdf

42. 2021 AWaRe classification [Internet]. Geneva: World Health Organization; 2021 Sep [cited 2022 Aug 8]. Report No.: WHO/HMP/HPS/EML/2021.04. Available from: https://www.who.int/publications-detail-redirect/2021-aware-classification

43. World Health Organization. 1 in 10 medical products in developing countries is substandard or falsified [Internet]. World Health Organization. 2017 [cited 2021 Dec 8]. Available from: https://www.who.int/news/item/28-11-2017-1-in-10-medical-products-in-developing-countries-is-substandard-or-falsified

44. World Health Organizaton. Index of World Pharmacopoeias and Pharmacopoeial authorities; QAS/11.453/Rev.14 [Internet]. WHO; 2022 [cited 2022 Oct 13]. Available from: https://www.who.int/publications/m/item/QAS-11.453-Rev.12

45. World Health Organisaton. Survey of the quality of selected antimalarial medicines circulating in six countries of sub-Saharan Africa [Internet]. Geneva: WHO; 2011 [cited 2016 Sep 5]. Available from: http://www.who.int/medicines/publications/qamsareport/en/

46. Singal GL, Nanda A, Kotwani A. A comparative evaluation of price and quality of some branded versus branded–generic medicines of the same manufacturer in India. Indian J Pharmacol. 2011 Apr;43(2):131–6.

47. Rahman MS, Yoshida N, Tsuboi H, Sokchamroeun U, Keila T, Sovannarith T, et al. A Cross-Sectional Investigation of the Quality of Selected Medicines for Noncommunicable Diseases in Private Community Drug Outlets in Cambodia during 2011–2013. The American Journal of Tropical Medicine and Hygiene. 2019 Nov 6;101(5):1018–26.

48. Jayabuana NN. Sistem Lelang JKN Berisiko Gerus Kualitas Obat. 2017 Nov 9 [cited 2018 Aug 6]; Available from: http://industri.bisnis.com/read/20171109/257/707814/sistem-lelang-jkn-berisiko-gerus-kualitas-obat

49. Sampurno. BPJS dan E-katalog Obat-obatan. Kompas [Internet]. 2018 Aug 7 [cited 2022 Oct 11]; Available from: https://www.kompas.id/baca/lain-lain/2018/08/07/bpjs-dan-e-katalog-obat-obatan-2

50. Hassali MA, Wong ZY, Alrasheedy AA, Saleem F, Mohamad Yahaya AH, Aljadhey H. Perspectives of physicians practicing in low and middle income countries towards generic medicines: A narrative review. Health Policy. 2014 Sep 1;117(3):297–310.

51. Patel A, Gauld R, Norris P, Rades T. “This body does not want free medicines”: South African consumer perceptions of drug quality. Health Policy and Planning. 2010 Jan 1;25(1):61–9.

52. Riner B, Bussy A, Hélène-Pelage J, Moueza N, Lamy S, Carrère P. ‘No generics, Doctor!’ The perspective of general practitioners in two French regions. BMC Health Serv Res. 2017 Nov 9;17(1):707.

53. Dunne SS, Dunne CP. What do people really think of generic medicines? A systematic review and critical appraisal of literature on stakeholder perceptions of generic drugs. BMC Medicine. 2015 Jul 29;13(1):173.

54. Pisani E, Dewi A, Palagyi A, Praveen D, Ihsan BRP, Lawuningtyas Hariadini A, et al. Variation in Price of Cardiovascular and Diabetes Medicine in Indonesia, and Relationship with Quality: A Mixed Methods Study in East Java. Am J Trop Med Hyg. 2023 Jun 7;108(6):1287–99.

55. Wagnild JM, Lee D, Jayeola B, Lukito PK, Fimbo A, Hampshire K. Can a Smartphone Application Help Address Barriers to Reporting Substandard/Falsified Medical Products? A Pilot Study in Tanzania and Indonesia. Glob Health Sci Pract. 2023 Aug 3;ghsp;GHSP-D-23-00034v1.

56. Vickers S, Bernier M, Zambrzycki S, Fernandez FM, Newton PN, Caillet C. Field detection devices for screening the quality of medicines: a systematic review. BMJ Global Health. 2018 Aug 1;3(4):e000725.

57. Waffo Tchounga CA, Sacré PY, Ravinetto R, Lieberman M, Hamuli Ciza P, Ngono Mballa R, et al. Usefulness of medicine screening tools in the frame of pharmaceutical post-marketing surveillance. PLoS One. 2023;18(8):e0289865.

58. Ching C, Zaman M h. Impact of ciprofloxacin impurities on bacterial growth, antibiotic resistance development and content assays. Letters in Applied Microbiology. 2021;73(2):220–8.